

ระบบการรวบรวมผลงานนักวิจัยจากฐานข้อมูลออนไลน์
แบบอัตโนมัติ
AUTOMATIC PUBLICATION CRAWLING SYSTEM



ชนากานต์ เจริญวิเชียรฉาย

สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2558

ระบบการรวบรวมผลงานนักวิจัยจากฐานข้อมูลออนไลน์
แบบอัตโนมัติ
AUTOMATIC PUBLICATION CRAWLING SYSTEM



สหกิจศึกษาเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

AUTOMATIC PUBLICATION CRAWLING SYSTEM



COOPERATIVE EDUCATION SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

ACADEMIC YEAR 2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา ระบบการรวบรวมผลงานนักวิจัยจากฐานข้อมูลออนไลน์แบบอัตโนมัติ
Automatic Publication Crawling System
ชื่อนักศึกษา นางสาวชนากานต์ เจริญวิเชียรฉาย 55050259
ปริญญา วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา วิทยาการคอมพิวเตอร์
ปีการศึกษา 2558
อาจารย์ที่ปรึกษา ผศ. กฤษฎา บุศรา

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต
(วิทยาการคอมพิวเตอร์) ประจำปีการศึกษา 2558

คณะกรรมการสอบ	ลายมือชื่อ
ผศ.กฤษฎา บุศรา ประธานกรรมการและอาจารย์ที่ปรึกษา	
ดร.ไพรัตน์ ธรเจริญศรี กรรมการ	

ลิขสิทธิ์ของคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	ระบบการรวบรวมผลงานนักวิจัยจากฐานข้อมูลออนไลน์แบบอัตโนมัติ Automatic Publication Crawling System
ชื่อนักศึกษา	นางสาวชนากานต์ เจริญวิเชียรฉาย 55050259
ปริญญา	วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2558
อาจารย์ที่ปรึกษา	ผศ. กฤษฎา บุศรา

บทคัดย่อ

เครือข่ายอินเทอร์เน็ตได้เข้ามามีบทบาทในการค้นหาข้อมูลต่างๆ เพิ่มมากขึ้นถือเป็นแหล่งข้อมูลออนไลน์ ที่มีความสำคัญต่อการค้นหาข้อมูลผลงานทางวิชาการเป็นอย่างยิ่ง เพื่อนำมาใช้เป็นฐานข้อมูลของระบบจัดลำดับผู้เชี่ยวชาญจากข้อมูลผลงานซึ่งระบบส่วนใหญ่ฐานข้อมูลไม่ได้ถูกปรับปรุง เมื่อข้อมูลผลงานมีการเปลี่ยนแปลงและเมื่อมีข้อมูลรายชื่อนักวิจัยใหม่เกิดขึ้น ระบบการรวบรวมผลงานนักวิจัยจาก แหล่งข้อมูลออนไลน์แบบอัตโนมัติ จะเป็นการรวบรวมข้อมูลผลงานวิจัยด้านวิทยาศาสตร์ของนักวิจัยในประเทศไทยแบบอัตโนมัติ เพื่อให้ข้อมูลผลงานที่ใช้ในการจัดลำดับผู้เชี่ยวชาญมีการปรับปรุงอยู่ตลอดเวลา ส่งผลให้การจัดลำดับผู้เชี่ยวชาญมีความถูกต้องมากขึ้น

คำสำคัญ : ระบบการรวบรวมผลงานนักวิจัย ระบบจัดลำดับผู้เชี่ยวชาญ

Title	Automatic Publication Crawling System
Student	Miss Chanakan Charoenvichenchai Student ID 55050259
Degree	Bachelor of Science Program (Computer Science)
Department	Computer Science
Faculty	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2015
Advisor	Asst.Prof.Krudsada Budsara

Abstract

Internet network has become a major part of people's lives including searching information. Internet is the key of researching. There are many experts whom can be used for reference in research. Come the question, who is the best in specific field. Now experts ranking become an issue. Old ranking used the relational databases which lack of updating data, so the ranking become outdated and useless. Because of that, new automatic collecting data system was invented. This system will automatically collect new information of experts from every online sources for improving Thai experts ranking. And now experts ranking become more accurate.

Keywords : Automatic collecting data system, Experts ranking

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงไปได้ด้วยความอนุเคราะห์อย่างยิ่ง ของ ดร.วสันต์ ภัทร อธิคม หัวหน้าห้องปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ที่ได้สละเวลาในการให้คำแนะนำ ชี้แนะให้เห็นถึงปัญหาและเสนอแนวทางในการแก้ปัญหาและขอบคุณยิ่ง นายณรงค์ อินทร์ธีรภัช ที่ได้คำแนะนำปรึกษาอย่างใกล้ชิด รวมทั้งพี่นักวิจัยจากห้อง ปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะ ที่ได้สละเวลาให้ความรู้ ให้คำปรึกษา ตรวจสอบ ให้คำแนะนำแนวทางการวิจัย และสนับสนุน จนทำให้การวิจัยในครั้งนี้สำเร็จออกมาด้วยดี ข้าพเจ้าจึงขอกราบระลึกถึงพระคุณไว้ ณ ที่นี้

นอกจากนี้การเข้าร่วมฝึกงานกับ ห้องปฏิบัติการวิจัยระบบขนส่งและจราจรอัจฉริยะดังกล่าว จะเกิดขึ้นไม่ได้ หากขาดคณะอาจารย์ที่ปรึกษา อันได้แก่ ผศ.กฤษฎา บุศรา ดร.ไพรัตน์ ธรเจริญศรี ผู้ซึ่งช่วยคอยให้คำแนะนำและความช่วยเหลือในด้านต่างๆ ที่เอื้อเพื่อให้การทำวิทยานิพนธ์ฉบับนี้สำเร็จ ลุล่วงไปด้วยดี

สุดท้ายนี้ ผู้เสนอวิทยานิพนธ์ขอกราบขอบพระคุณบิดามารดา และครอบครัว สำหรับกำลังใจ ที่มีค่ายิ่ง รวมถึงขอขอบพระคุณเพื่อนร่วมงาน และมิตรสหาย ที่คอยติดตามให้กำลังใจ ให้การสนับสนุนและความช่วยเหลือในด้านต่างๆ และท่านอื่นๆ ที่มีได้กล่าวชื่อไว้ ณ ที่นี้ที่มีส่วนช่วยให้ วิทยานิพนธ์ของข้าพเจ้าสำเร็จไปได้ด้วยดี

ชนากานต์ เจริญวิเชียรฉาย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ฉ
สารบัญรูป.....	ช
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	1
1.3 ขอบเขตของงานวิจัย.....	1
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	3
2.1 ภาษาที่ใช้ในการพัฒนา.....	3
2.1.1 PHP.....	3
2.1.2 SQL.....	3
2.2 ทฤษฎีที่เกี่ยวข้อง.....	3
2.2.1 การเชื่อมต่อข้อมูลแบบ API.....	3
2.2.2 Text Classification.....	6
2.2.3 Stop Words.....	6
2.2.3 Text Mining.....	6
2.3 โปรแกรมที่ใช้.....	7
2.3.1 Notepad++.....	7
2.3.2 Putty.....	7
2.4 งานวิจัยที่เกี่ยวข้อง.....	7
2.4.1 Remembrance Agent: A Continuously Running Automated Information Retrieval System.....	7
2.4.2 Expertise Ranking For Online Social Network Group Members.....	7
บทที่ 3 วิธีการดำเนินงานวิจัย.....	8
3.1 สืบค้นเว็บไซต์ที่ให้บริการ API.....	9
3.2 ดึงข้อมูลจากเว็บไซต์ผ่าน API.....	10

สารบัญ (ต่อ)

	หน้า
3.2.1 IEEE Xplore.....	10
3.2.2 Scopus	11
3.2.3 PubMed	14
3.2.4 SlideShare	18
3.2.5 Google News Search	23
3.2.6 Google Video Search.....	24
3.2.7 Google Books APIs.....	25
3.2.8 Inspire-Hep.....	32
3.2.9 DBLP	33
3.2.10 Arxiv.....	36
3.3 แปลงข้อมูลจาก XML/JSON ให้อยู่ในรูปแบบที่ต้องการ.....	37
3.3.1 ขั้นตอนการแปลงข้อมูล XML.....	37
3.3.2 ขั้นตอนการแปลงข้อมูล JSON.....	38
3.4 ทำการ Mapping ข้อมูลผลงานกับสาขา.....	39
3.5 เก็บลงฐานข้อมูล.....	40
บทที่ 4 ผลการวิจัยและการอภิปราย.....	42
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	44
5.1 สรุปผลการวิจัย.....	44
5.2 ข้อเสนอแนะ.....	44
เอกสารอ้างอิง	45

สารบัญตาราง

ตารางที่	หน้า
3.1 พารามิเตอร์ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ SlideShare.....	18
3.2 จำนวนผลงานที่ทำการรวบรวมข้อมูลจากเว็บไซต์ต่างๆ.....	41
4.1 จำนวนผลงาน และนักวิจัยที่มีผลงานจากเว็บไซต์ต่างๆ.....	43
4.2 จำนวนผลงาน และนักวิจัยที่มีผลงานตามประเภทผลงาน.....	43



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างของภาษา XML	4
2.2 โครงสร้าง Object ของ JSON	5
2.3 โครงสร้าง Array ของ JSON	5
2.4 ตัวอย่างของภาษา JSON	5
2.5 ตัวอย่างของของการทำ Text Classification	6
3.1 ภาพรวมแนวทางในการดำเนินงาน	8
3.2 โลโก้เว็บไซต์ที่ทำการสืบค้น	9
3.3 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ IEEE Xplore	10
3.4 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ IEEE Xplore	10
3.5 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ IEEE Xplore.....	11
3.6 หน้าเว็บไซต์ “ELSEVIER”	12
3.7 หน้าสมัครสมาชิกของ เว็บไซต์ “ELSEVIER”	12
3.8 การสมัครเพื่อขอ “API Key” ของเว็บไซต์ Scopus	12
3.9 การสร้าง API Key ของเว็บไซต์ Scopus	13
3.10 วิธีการดู “API Key” ที่ใช้ในการดึงข้อมูลของเว็บไซต์ Scopus	13
3.11 “API Key” ที่ใช้ในการดึงข้อมูลของเว็บไซต์ Scopus	13
3.12 ตัวอย่าง URL ที่ใช้ในการดึงข้อมูลผ่าน API ของเว็บไซต์ Scopus	13
3.13 ข้อมูลผลงานเบื้องต้นของนักวิจัยของเว็บไซต์ Scopus	14
3.14 ตัวอย่าง URL ที่ใช้ในการดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ Scopus	14
3.15 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Scopus	14
3.16 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ PubMed	15
3.17 ตัวอย่าง URL ที่ใช้ดึงข้อมูลไอทีผลงานผ่าน API ของเว็บไซต์ PubMed	15
3.18 ข้อมูลรายการไอทีผลงานของนักวิจัยในรูปแบบของ JSON ของเว็บไซต์ PubMed.....	16
3.19 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ PubMed.....	16
3.20 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ PubMed	17
3.21 ตัวอย่าง URL ที่ใช้ดึงข้อมูล “Cited” ของผลงานของเว็บไซต์ PubMed.....	17
3.22 ตัวอย่าง Source Code ของหน้าเว็บไซต์ PubMed	17
3.23 หน้าเว็บไซต์ SlideShare	18
3.24 การสมัครสมาชิกของเว็บไซต์ SlideShare	19
3.25 การเข้าถึงหน้าเว็บไซต์ที่แสดงรายละเอียดการดึงข้อมูลผ่าน API ของเว็บไซต์ SlideShare	19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.26 วิธีการเข้าถึงหน้าเว็บไซต์ที่ขอ API Key ของเว็บไซต์ SlideShare.....	19
3.27 หน้ากรอกคำขอ API Key ของเว็บไซต์ SlideShare.....	20
3.28 “API key” และ “Shared secret”ของเว็บไซต์ SlideShare.....	20
3.29 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ SlideShare.....	21
3.30 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ SlideShare	21
3.31 ตัวอย่าง URL ที่ใช้ในการดึงข้อมูลของเว็บไซต์ SlideShare	21
3.32 หน้าเว็บไซต์ผลงานในส่วนของคุณสมบัติของข้อมูล Statistics ของเว็บไซต์ SlideShare	22
3.33 ข้อมูล Statistics ของผลงานในรูปแบบของ XML ของเว็บไซต์ SlideShare.....	22
3.34 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search.....	23
3.35 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search	23
3.36 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Google News Search.....	24
3.37 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ Google Video Search	24
3.38 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search.....	25
3.39 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Google Video Search	25
3.40 หน้าเว็บไซต์ Google Books APIs.....	26
3.41 หน้าลงชื่อเข้าใช้ของเว็บไซต์ Google Books APIs.....	26
3.42 วิธีการเข้าหน้า Google Developers Console.....	26
3.43 วิธีการเข้าหน้าเว็บไซต์ที่ใช้สร้างโปรเจกต์ของเว็บไซต์ Google Developers Console	27
3.44 วิธีการสร้างโปรเจกต์ของเว็บไซต์ Google Developers Console	27
3.45 วิธีการเข้าหน้าเว็บไซต์ที่ใช้จัดการ API ของเว็บไซต์ Google Developers Console	28
3.46 การเลือกเมนู “Credentials” ของเว็บไซต์ Google Developers Console	28
3.47 การเพิ่ม “Credentials” ของเว็บไซต์ Google Developers Console	29
3.48 การสร้าง API Key ของเว็บไซต์ Google Developers Console.....	29
3.49 API Key ของเว็บไซต์ Google Developer Console	30
3.50 การเลือกเมนู “Overview” ของเว็บไซต์ Google Developer Console	30
3.51 การค้นหา Books API ของเว็บไซต์ Google Developer Console	31
3.52 การทำให้ API Key ของโปรเจกต์สามารถใช้งาน Books API ได้.....	31
3.53 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google Books APIs	31
3.54 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Google Books APIs.....	32
3.55 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ Inspirehep	32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.56 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ Inspirehep.....	33
3.57 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Inspirehep.....	33
3.58 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ DBLP.....	34
3.59 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ DBLP	34
3.60 ชื่อนักวิจัยตามรูปแบบที่เว็บไซต์กำหนดในรูปแบบ XML ของเว็บไซต์ DBLP.....	34
3.61 ตัวอย่าง URL ที่ใช้ดึงรายการไอดีผลงานของนักวิจัยผ่าน API ของเว็บไซต์ DBLP	34
3.62 รายการไอดีผลงานของนักวิจัย (dblpkey) ในรูปแบบของ XML ของเว็บไซต์ DBLP.....	35
3.63 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานของนักวิจัยผ่าน API ของเว็บไซต์ DBLP	35
3.64 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ DBLP.....	35
3.65 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ Arxiv.....	36
3.66 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ Arxiv.....	36
3.67 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ Arxiv.....	37
3.68 ตัวอย่างไฟล์ XML เรียกว่า "books.xml".....	37
3.69 ตัวอย่างคำสั่งในการแปลงข้อมูล XML.....	38
3.70 ตัวอย่างข้อมูลที่ได้จากการแปลงข้อมูล XML.....	38
3.71 ตัวอย่างไฟล์ JSON ที่เรียกว่า "employees.json".....	38
3.72 ตัวอย่างคำสั่งในการแปลงข้อมูล JSON	39
3.73 ตัวอย่างข้อมูลที่ได้จากการแปลงข้อมูล JSON.....	39
3.74 ขั้นตอนการ Mapping ข้อมูลผลงานกับสาขา.....	40
3.75 ตัวอย่างข้อมูลที่ทำให้การบันทึกลงฐานข้อมูล	40

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเครือข่ายอินเทอร์เน็ตได้เข้ามามีบทบาทในการดำเนินชีวิตเพิ่มมากขึ้น ซึ่งมีอิทธิพลและส่งผลกระทบต่อปัจจัยในด้านการทำงาน การติดต่อสื่อสาร การเรียนการสอน แหล่งความรู้ต่างๆ สิ่งเหล่านี้ ล้วนอยู่บนอินเทอร์เน็ตทั้งสิ้น ทำให้การศึกษามีความก้าวหน้าไปอย่างรวดเร็ว และส่งเสริมการค้นคว้า และค้นหาข้อมูลงานวิจัยและองค์ความรู้ใหม่ๆ ที่สำคัญทางด้านการศึกษา และด้านธุรกิจ โดยเฉพาะอย่างยิ่งในวงการการศึกษาและธุรกิจที่เกี่ยวข้องกับการวิจัย การค้นหาผู้ที่มีความเชี่ยวชาญในด้านต่างได้นั้นถือเป็นเรื่องยาก เพราะการค้นหาข้อมูลในปัจจุบันยังให้ผลการค้นหาแบบวงกว้าง ไม่ได้มีการเจาะลึกถึงรายละเอียดมากนัก ทำให้เมื่อผู้สนใจที่ต้องการจะทราบชื่อนักวิทยาศาสตร์ที่มีความเชี่ยวชาญในสาขาใดๆ (อาทิเช่น นักเรียนและนักศึกษาต้องการขอคำปรึกษาและแต่งตั้งที่ปรึกษาจากผู้เชี่ยวชาญ หรือการค้นหาผู้เชี่ยวชาญ สำหรับการให้คำปรึกษาทางเฉพาะทางโดยภาครัฐหรือภาคเอกชน เป็นต้น) ไม่สามารถที่จะค้นหาข้อมูลนักวิจัยได้อย่างเหมาะสม เนื่องจากฐานข้อมูลนักวิทยาศาสตร์ในประเทศไทยยังไม่สมบูรณ์ มีข้อมูลผลงานวิจัยไม่เพียงพอการจัดกระจาย และฐานข้อมูลไม่ได้ถูกปรับปรุง เมื่อมีการเปลี่ยนแปลงผลงาน และรายชื่อนักวิจัยเกิดขึ้น

ดังนั้น มีความจำเป็นต้องจัดทำ ระบบการรวบรวมผลงานนักวิจัยจากแหล่งข้อมูลออนไลน์แบบอัตโนมัติ เพื่อทำให้ข้อมูลผลงานที่ใช้ในการจัดลำดับผู้เชี่ยวชาญมีการปรับปรุงอยู่ตลอดเวลา ส่งผลให้การจัดลำดับผู้เชี่ยวชาญมีความถูกต้องมากขึ้น

1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อรวบรวมข้อมูลผลงานวิจัยด้านวิทยาศาสตร์ของนักวิจัยในประเทศไทยแบบอัตโนมัติ
- 2) เพื่อให้ฐานข้อมูลผลงานวิจัยมีการปรับปรุงอยู่ตลอดเวลา

1.3 ขอบเขตของงานวิจัย

งานวิจัยนี้เป็นการวิจัยเกี่ยวกับการรวบรวมจัดลำดับผู้เชี่ยวชาญมีขอบเขตในการวิจัยดังนี้

- 1) ข้อมูลนักวิทยาศาสตร์ภายในกระทรวงวิทยาศาสตร์
- 2) ข้อมูลผลงานนักวิจัยดึงมาจากเครือข่ายสังคม (Social Network)
 - ฐานข้อมูลวิจัยออนไลน์ เช่น IEEE, Scopus, Pubmed
 - สื่อสังคมออนไลน์ (social media) เช่น Youtube, twitter
 - ข้อมูลออนไลน์อื่นๆ เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- งานนำเสนอ (Slideshare)
- ข่าวงานวิจัย (Google News Search)
- หนังสือทางวิชาการ (Google Books APIs)
- การบรรยายทางวิชาการ (Google Video search)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ข้อมูลผลงานวิจัยด้านวิทยาศาสตร์ภายในประเทศไทยมีความทันสมัยเป็นปัจจุบัน
- 2) สามารถจัดลำดับผู้เชี่ยวชาญในสาขาต่างๆ มีความแม่นยำมากและเป็นปัจจุบันมากขึ้น การจัดลำดับนักวิจัยมีการปรับปรุงอยู่ตลอดเวลา



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ภาษาที่ใช้ในการพัฒนา

2.1.1 PHP

PHP เป็นภาษาคอมพิวเตอร์ประเภทโอเพนซอร์สสำหรับพัฒนาเว็บเพจเมื่อเครื่องบริการได้รับคำสั่งจากผู้ใช้ก็จะส่งให้กับตัวแปลภาษา ทำหน้าที่ประมวลผลและส่งข้อมูลกลับไปยังเครื่องของผู้ใช้ที่ร้องขอ ในรูป HTML ภาพ หรือแฟ้ม Digital อื่นๆ ลักษณะของภาษามีคำสั่งมาจากภาษาซี ภาษาจาวาและภาษาเพิร์ล ซึ่ง ภาษา PHP นั้นง่ายต่อการเรียนรู้ ซึ่งเป้าหมายหลักของภาษานี้คือให้นักพัฒนาเว็บไซต์สามารถเขียนเว็บเพจที่มีความตอบโต้ ได้อย่างรวดเร็ว PHP เป็นภาษาจำพวกภาษาสคริปต์คำสั่งต่างๆจะเก็บอยู่ในไฟล์ที่เรียกว่า สคริปต์ และเวลาใช้งานต้องอาศัยตัวแปลชุดคำสั่ง ตัวอย่างของภาษาสคริปต์ เช่น JavaScript, Perl เป็นต้น ลักษณะของ PHP ที่แตกต่างจากภาษาสคริปต์แบบอื่นๆ คือ PHP ได้รับการพัฒนาและออกแบบมาเพื่อ ใช้งานในการสร้างเอกสารแบบ HTML โดยสามารถสอดแทรก หรือแก้ไขเนื้อหาได้โดยอัตโนมัติ [1]

2.1.2 SQL

SQL ย่อมาจาก Structured Query Language คือ ภาษาที่ใช้ในการเขียนโปรแกรม เพื่อจัดการกับฐานข้อมูลโดยเฉพาะ เป็นภาษามาตรฐานบนระบบฐานข้อมูลเชิงสัมพันธ์ และเป็นระบบเปิด (open system) หมายถึงเราสามารถใส่คำสั่ง SQL กับฐานข้อมูลชนิดใดก็ได้ และคำสั่งงานเดียวกันเมื่อสั่งงานผ่านระบบฐานข้อมูลที่แตกต่างกันจะได้ผลลัพธ์เหมือนกัน ทำให้เราสามารถเลือกใช้ฐานข้อมูล ชนิดใดก็ได้โดยไม่ติดขัดกับฐานข้อมูลใดฐานข้อมูลหนึ่ง นอกจากนี้แล้ว SQL ยังเป็นชื่อโปรแกรมฐานข้อมูล ซึ่งโปรแกรม SQL เป็นโปรแกรมฐานข้อมูลที่มีโครงสร้างของภาษาที่เข้าใจง่าย ไม่ซับซ้อน มีประสิทธิภาพการทำงานสูง สามารถทำงานที่ซับซ้อนได้โดยใช้คำสั่งเพียงไม่กี่คำสั่ง โปรแกรม SQL จึงเหมาะที่จะใช้กับระบบฐานข้อมูลเชิงสัมพันธ์ ซึ่งแบ่งการทำงานได้เป็น 4 ประเภท ดังนี้ Select query ใช้สำหรับดึงข้อมูลที่ต้องการ, Update query ใช้สำหรับแก้ไขข้อมูล, Insert query ใช้สำหรับการเพิ่มข้อมูล, Delete query ใช้สำหรับลบข้อมูลออกไป [2]

2.2 ทฤษฎีที่เกี่ยวข้อง

2.2.1 การเชื่อมต่อข้อมูลแบบ API

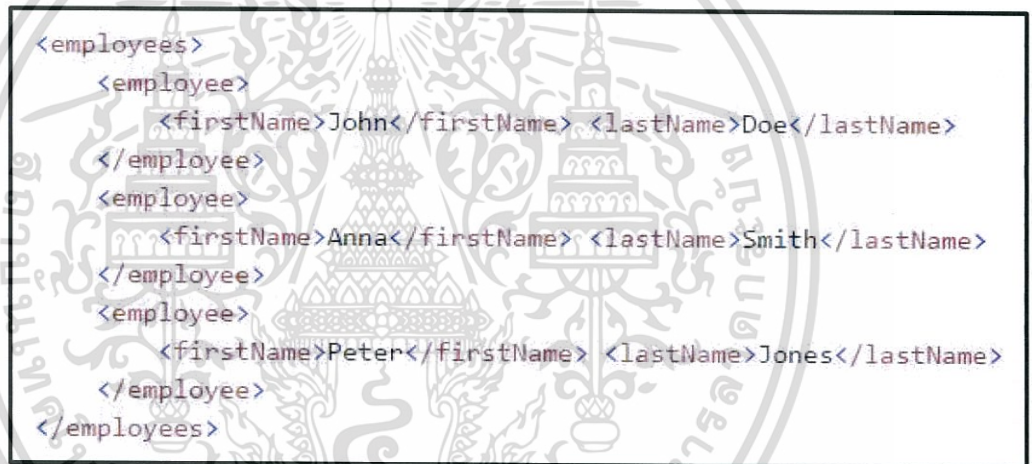
API ย่อมาจาก Application Programming Interface เป็นภาษาคอมพิวเตอร์ ที่ทำให้คอมพิวเตอร์ติดต่อพูดคุยกันเองได้โดย API เป็นช่องทางการในการเชื่อมต่อกับเว็บไซต์ เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของศูนย์วิจัยและพัฒนาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้ให้บริการ API ทำหน้าที่ช่วยในการเข้าถึงข้อมูลต่างภายในเว็บไซต์ ตามที่ผู้ให้บริการเว็บไซต์ ที่มี API กำหนดขอบเขตในการเข้าถึงข้อมูล บนเว็บไซต์นั้น โดยรูปแบบข้อมูลที่ใช้ส่งผ่าน API มี 2 รูปแบบที่นิยมใช้ในปัจจุบันได้แก่ XML และ JSON

1) รูปแบบการส่งข้อมูลแบบ XML

XML ย่อมาจาก Extensible Markup Language คือ ภาษาที่ใช้ในการแสดงผล โดยจะทำการเก็บข้อมูลและโครงสร้างเข้าไว้ด้วยกัน โดยภาษา XML มีโครงสร้างที่ประกอบด้วยแท็ก (Tag) เปิดตามด้วยข้อมูลและแท็ก (Tag) ปิดเช่นเดียวกับภาษา HTML แต่ภาษา XML จะสามารถสร้างแท็ก กำหนดชื่อแท็ก ชื่อแอตทริบิวต์ (Element) และกำหนดโครงสร้างของข้อมูลได้เอง [3] โดย XML ประกอบไปด้วย

- แท็ก (Tag)
- ข้อมูล (Data)



```
<employees>
  <employee>
    <firstName>John</firstName> <lastName>Doe</lastName>
  </employee>
  <employee>
    <firstName>Anna</firstName> <lastName>Smith</lastName>
  </employee>
  <employee>
    <firstName>Peter</firstName> <lastName>Jones</lastName>
  </employee>
</employees>
```

รูปที่ 2.1 ตัวอย่างของภาษา XML

จากรูปที่ 2.1 จะแสดงตัวอย่างข้อมูลที่ถูกเก็บอยู่ในรูปแบบ XML โดยข้อมูล “John” จะถูกเก็บอยู่ในแท็กชื่อ firstName และข้อมูล “Doe” จะถูกเก็บอยู่ในแท็กชื่อ lastName โดยที่แท็ก firstName และ lastName จะอยู่ในแท็กชื่อ employee และแท็ก employee จะถูกเก็บในแท็กชื่อ employees อีกชั้นหนึ่ง [7]

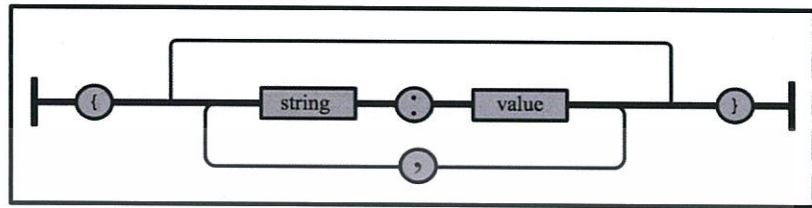
2) รูปแบบการส่งข้อมูลแบบ JSON

JSON ย่อมาจาก JavaScript Object Notation เป็นรูปแบบภาษาที่ใช้ในการแลกเปลี่ยนข้อมูล ซึ่งมีโครงสร้างและรูปแบบที่ง่าย สะดวก และรวดเร็วกว่า XML โดย JSON ส่วนใหญ่นำมาใช้ในการทำ AJAX เพื่อทำให้สามารถรับส่งค่า หรือสั่งการเข้า Server พร้อมกับชุดข้อมูลขนาดใหญ่ได้รูปแบบของ JSON จะมีอยู่สองประเภทด้วยกัน คือ Object และ Array โดยรูปแบบของ JSON มีดังนี้

1) Object

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

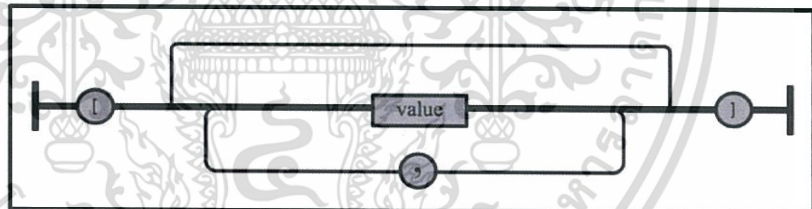
Object คือ ชุดไม่เรียงลำดับและไม่ซ้ำกัน (Unordered Set) ของคู่ Key/Value ซึ่งรูปแบบของ Object ใน JSON จะเป็นไปตามรูปที่ 2.2 โดยจะเริ่มต้นด้วย “{” (Left Brace) และจบด้วย “}” (Right Brace) โดย Key แต่ละตัวจะถูกตามด้วย “:” (Colon) และ คู่ Key/Value แต่ละคู่จะถูกคั่นด้วย “,” (Comma)



รูปที่ 2.2 โครงสร้าง Object ของ JSON

2) Array

Array คือ ชุดที่เรียงลำดับและซ้ำกันได้ (Ordered Collection) ของ Value ซึ่งรูปแบบของ Array ใน JSON จะเป็นไปตามรูปที่ 2.3 โดยจะเริ่มต้นด้วย “[” (Left Bracket) และจบด้วย “]” (Right Bracket) โดย Value แต่ละตัวจะถูกคั่นด้วย “,” (Comma) [4]



รูปที่ 2.3 โครงสร้าง Array ของ JSON

```
{
  "employees": [
    {
      "firstName": "John",
      "lastName": "Doe"
    },
    {
      "firstName": "Anna",
      "lastName": "Smith"
    },
    {
      "firstName": "Peter",
      "lastName": "Jones"
    }
  ]
}
```

รูปที่ 2.4 ตัวอย่างของภาษา JSON

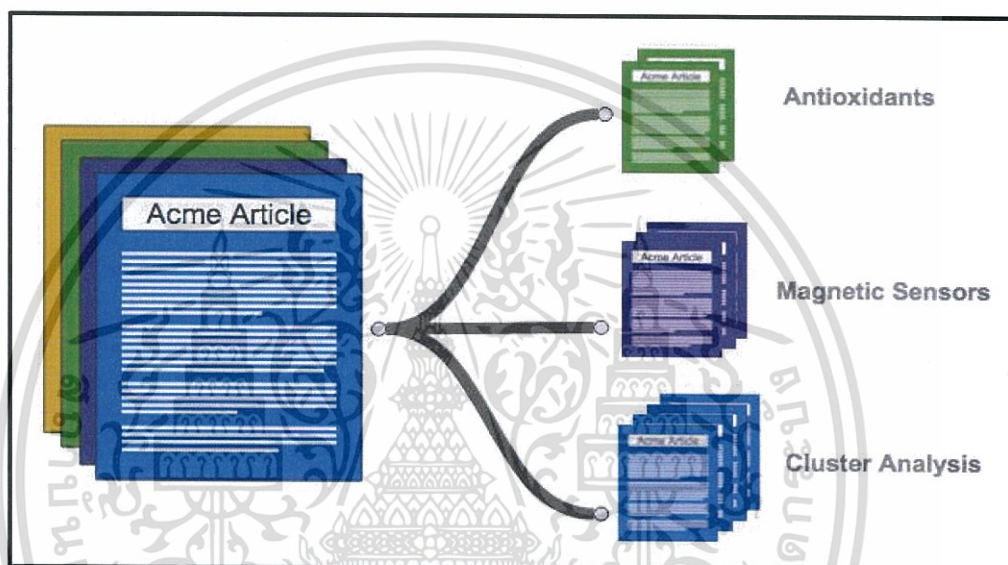
จากรูปที่ 2.4 ตัวอย่างของภาษา JSON จะแสดง Object ที่มี Key ชื่อ “employees” และ Value ของ Key “employees” คือ Array ที่มี Value เป็น Object ซึ่งแต่ละ Object จะมี Key ชื่อ firstName และ lastName

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่แต่ละ Key จะมี Value อีกที เช่น ข้อมูล “John” จะถูกเก็บอยู่ใน key ชื่อ firstName และข้อมูล “Doe” จะถูกเก็บอยู่ใน key ชื่อ lastName [5]

2.2.2 Text Classification

Text Classification คือ การหากฎเพื่อระบุประเภทของข้อมูลจากคุณสมบัติของข้อมูลหรือ การหาความสัมพันธ์ระหว่างข้อมูล เช่น การจับคู่โดยการค้นหาคำหลักในข้อมูล ผลงานที่เราต้องการค้นหาถ้าพบคำหลักก็ทำการจัดประเภทข้อมูลนั้นให้อยู่ในประเภทคำหลักนั้น ตามรูปที่ 2.5



รูปที่ 2.5 ตัวอย่างของการทำ Text Classification

2.2.3 Stop Words

Stop Words คือ คำที่ไม่มีความหมายต่อเนื้อหาในเอกสารนั้นๆ และสามารถตัดทิ้งได้ การตัดคำที่ไม่สำคัญทิ้งจะไม่มีผลต่อการทำดัชนี (index) และไม่ทำให้ประสิทธิภาพการค้นหาข้อมูลต่ำลง ตัวอย่าง คำที่ไม่สำคัญในภาษาอังกฤษ เช่น a, an, the, in, out, before, up, down เป็นต้น เนื่องจากคำที่ไม่สำคัญในภาษาอังกฤษมีการระบุไว้อย่างชัดเจนว่า มีคำอะไรบ้างในรายการ (Stop Word List) โดยสามารถแก้ไขและเปลี่ยนแปลงรายการได้ [6]

2.2.4 Text mining

Text mining คือ กระบวนการวิเคราะห์สารสนเทศ เพื่อค้นหารูปแบบแนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความนั้น ด้วยวิธีการสกัดคำ/ตัดคำ (Information Extraction) และ วิธีการทางสถิติ ด้วยโปรแกรมคอมพิวเตอร์แบบอัตโนมัตินำเสนอผลการวิเคราะห์ให้เป็นความรู้ใหม่ รวมถึงสามารถแสดงความสัมพันธ์ของข้อมูลใหม่ด้วย [7]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 โปรแกรมที่ใช้

2.3.1 Notepad++

Notepad++ เป็นโปรแกรม Text-Editor ที่รองรับภาษาเขียนโปรแกรมมากมาย ไม่ว่าจะเป็น C, C++, HTML, ASP, Java, Pascal, CSS และอื่นๆ อีกมากมาย พร้อมการไฮไลต์เปลี่ยนสีคำสั่ง (Syntax)

2.3.2 Putty

Putty เป็นโปรแกรม Telnet เอาไว้เชื่อมต่อข้อมูล ส่งชุดคำสั่งแบบ Command Line จากเครื่องลูก Client เพื่อที่จะเข้าไปจัดการในเครื่องแม่ Server ขนาดเล็ก

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 REMEMBRANCE AGENT: A CONTINUOUSLY RUNNING AUTOMATED INFORMATION RETRIEVAL SYSTEM

Remembrance Agent (RA) เป็นโปรแกรมซึ่งช่วยในเรื่องความจำของมนุษย์โดยการแสดงเป็นรายการเอกสาร ซึ่งระบบการดึงข้อมูลส่วนมากจะต้องให้ผู้ใช้งานทำการบันทึกสิ่งเหล่านั้นเอง แต่ RA จะทำการบันทึกแบบอัตโนมัติโดยผู้ใช้งานไม่ต้องมาทำการบันทึกเอง ผู้ใช้งานเพียงแต่ทำการเลือกที่จะติดตามหรือไม่ติดตาม ข้อเสนอแนะของระบบ RA ตามที่ต้องการ เช่น วิศวกรได้อ่านอีเมลที่เกี่ยวข้องกับโครงการหนึ่ง RA จะทำการเตือนเมื่อวิศวกรหยุดอ่านอีเมลและเริ่มที่จะทำงานอื่นๆ โดย RA จะทำการรวบรวมสิ่งที่เกี่ยวข้องกับโครงการนั้นไม่ว่าจะเป็นตารางโครงการ สถานะโครงการและทรัพยากรต่างๆที่เกี่ยวข้องจะทำการเปลี่ยนให้อยู่ในรูปแบบคำแนะนำหรือแสดงผลในบรรทัดเดียวเพื่อให้งานต่อการจดจำ [8]

2.4.2 EXPERTISE RANKING FOR ONLINE SOCIAL NETWORK GROUP MEMBERS

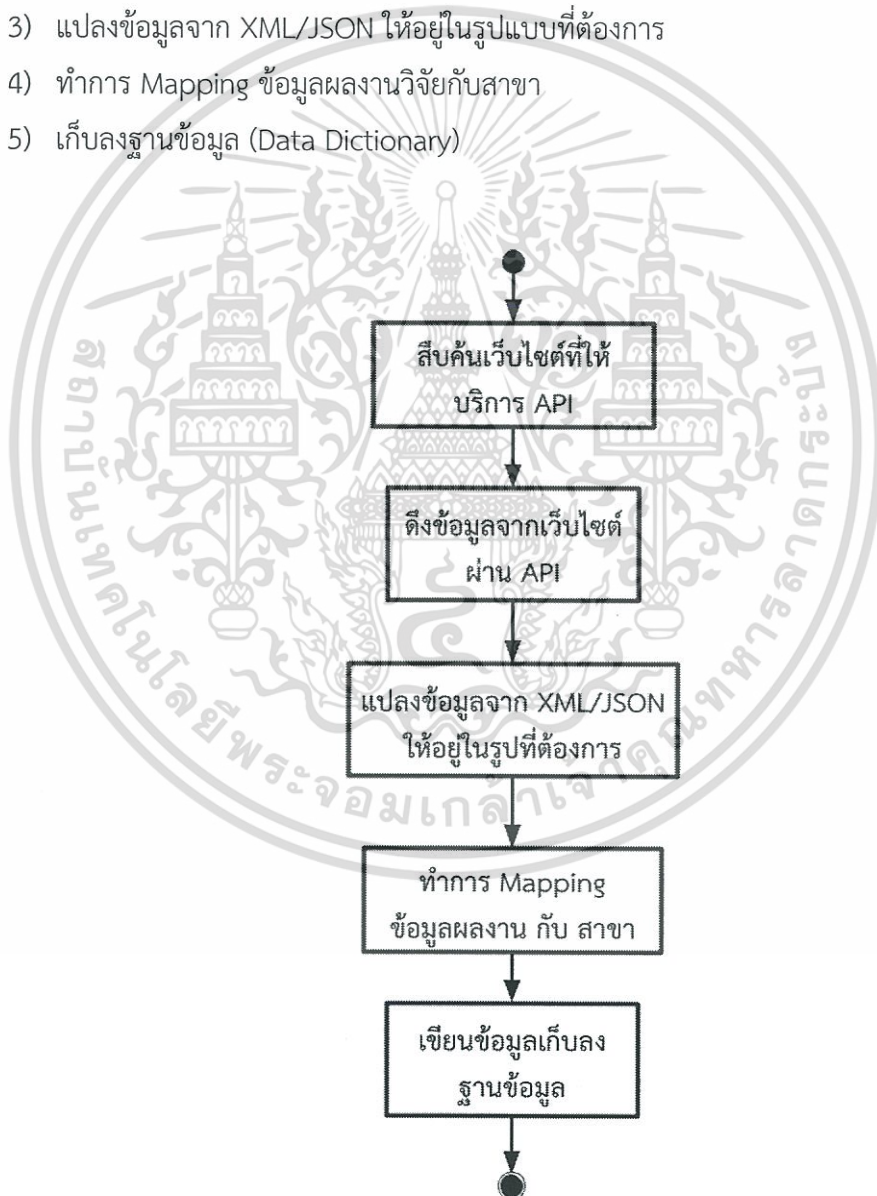
เป็นงานวิจัยเป็นการวิจัยเกี่ยวกับการทดลองการวัดประสิทธิภาพการจัดลำดับผู้เชี่ยวชาญถ้าพิจารณาความคิดเห็นภายในกลุ่มเฟซบุ๊กจาวาด้วย จะส่งผลต่อลำดับผู้เชี่ยวชาญแบบเพจแรงค์อย่างเดี่ยวหรือไม่ ซึ่งผลการทดลอง แม้ว่าการพิจารณาความคิดเห็นด้วยนั้นจะยังไม่ส่งผลเด่นชัดต่อการจัดลำดับความเชี่ยวชาญเนื่องด้วยข้อจำกัดของข้อมูลทดสอบและการประเมินผลแต่มีแนวโน้มที่จะส่งผลดีต่อการจัดลำดับความเชี่ยวชาญ [9]

บทที่ 3

วิธีการดำเนินงานวิจัย

งานวิจัยนี้จะนำเสนอกระบวนการในการรวบรวมผลงานนักวิจัยจาก ฐานข้อมูลออนไลน์แบบอัตโนมัติ โดยภาพรวมของขั้นตอนการดำเนินงานนั้นเป็นไปตามรูปที่ 3.1 ซึ่งประกอบด้วยขั้นตอน 5 ขั้นตอนคือ

- 1) สืบค้นเว็บไซต์ที่ให้บริการ API
- 2) ดึงข้อมูลจากเว็บไซต์ผ่าน API
- 3) แปลงข้อมูลจาก XML/JSON ให้อยู่ในรูปแบบที่ต้องการ
- 4) ทำการ Mapping ข้อมูลผลงานวิจัยกับสาขา
- 5) เก็บลงฐานข้อมูล (Data Dictionary)



รูปที่ 3.1 ภาพรวมแนวทางในการดำเนินงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1 สืบค้นเว็บไซต์ที่ให้บริการ API

การสืบค้นเว็บไซต์ทางวิชาการต่างที่ให้บริการ API จะเริ่มจากการค้นหารายชื่อเว็บไซต์ต่างที่ให้บริการ API ผ่านทาง “Search Engine” หรือเว็บไซต์ที่เราต้องการดึงข้อมูล เมื่อได้เว็บไซต์ที่ต้องการแล้ว ทำการศึกษารายละเอียดการให้บริการต่างๆของเว็บไซต์ว่ามีข้อจำกัดของการดึงข้อมูลอย่างไร เช่นอนุญาตให้ดึงข้อมูล 1000 ครั้งต่อวัน เป็นต้น ซึ่งเว็บไซต์ที่ทำการสืบค้นมามีดังนี้

- 1) IEEE Xplore
- 2) Scopus
- 3) PubMed
- 4) Slide Share
- 5) Google News Search
- 6) Google Books APIs
- 7) Google Video Search
- 8) Inspire-Hep
- 9) Arxiv
- 10) DBLP



รูปที่ 3.2 โลโก้เว็บไซต์ที่ทำการสืบค้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ดึงข้อมูลจากเว็บไซต์ผ่าน API

การดึงข้อมูลจากเว็บไซต์ผ่าน API ของแต่ละเว็บไซต์มีวิธีการดึงที่แตกต่างกันออกไป ซึ่งเป็นไปตามข้อกำหนดของเว็บไซต์นั้นๆ โดยการค้นหาข้อมูลต่างๆบนเว็บไซต์จะทำการค้นหาโดยใช้ชื่อนักวิจัยในการค้นหา

3.2.1 IEEE Xplore

IEEE Xplore เป็นฐานข้อมูลทางด้านวิศวกรรมไฟฟ้า อิเล็กทรอนิกส์และเทคโนโลยีสารสนเทศ ของ the Institute of Electrical and Electronics Engineers (IEEE) รวบรวมข้อมูลจากเอกสารการประชุมสัมมนา เอกสารมาตรฐาน และวารสารนิตยสาร ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ IEEE Xplore มีดังนี้

- 1) ใส่ URL: <http://ieeexplore.ieee.org/gateway> ที่เว็บเบราว์เซอร์จะปรากฏหน้าวิธีการ และ ตัวอย่างการดึงข้อมูลผ่าน API ของเว็บไซต์ “IEEE Xplore” [10] ตามรูปที่ 3.3

Query Parameters:		
The following parameters are available for querying the IEEE Xplore Search Gateway.		
Parameter	Description	Boolean query field
Parameter specifying single article		
an	Article number	'Article Number'
Parameters specifying search fields and terms		
au	Terms to search for in <i>Author</i>	'Author'
ti	Terms to search for in <i>Document Title</i>	'Document Title'
ab	Terms to search for in <i>Abstract</i>	'Abstract'
doi	Terms to search for in <i>DOI</i>	'DOI'
af	Terms to search for in <i>Affiliation</i>	'Author Affiliation'
pt	Terms to search for in <i>Publication Title</i>	'Publication Title'
isbn	Terms to search for in <i>ISBN</i>	'isbn'
issn	Terms to search for in <i>ISSN</i>	'issn'
py	Terms to search for in <i>Publication Year</i>	'Publication Year'
partnum	Terms to search by <i>Part Number</i>	'Part Number'
thstterms	Terms to search for in <i>Theaurus Terms</i>	'Thesaurus Terms'
cntrlterms	Terms to search for in <i>Controlled Index Terms</i>	'Inspec Controlled Terms'
idxterms	Terms to search for in <i>Index Terms</i>	'Search Index Terms'
and	Terms to search for in all configured metadata fields and abstract. Accepts complex queries involving field names and boolean operators.	
querytext	Terms to search for in all configured metadata fields, abstract and document text. Accepts complex queries involving field names and boolean operators.	
Filtering parameters		
oa	Open Access only (1 - true, 0 - false)	
pn	Publication number	
pys	Start value of <i>Publication Year</i> to restrict results by.	

รูปที่ 3.3 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ IEEE Xplore

- 2) จากนั้นทำการแทน ชื่อ-นามสกุล ของนักวิจัยลงใน URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ IEEE Xplore ตามรูปที่ 3.4 โดยใช้ชื่อ “Wasan Pattara-atikom” เป็นตัวอย่างในการค้นหาข้อมูล

URL = <http://ieeexplore.ieee.org/gateway/ipsSearch.jsp?querytext=java&au=Wasan+Pattara-atikom>

รูปที่ 3.4 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ IEEE Xplore

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของลิขสิทธิ์ หากมีข้อผิดพลาดประการใดขออภัยเป็นอย่างสูง และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 3) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบของ XML ตามรูปที่ 3.5 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ title, author, cited, publish year, keywords

```

<root>
  <totalfound>2</totalfound>
  <totalsearched>3814084</totalsearched>
  <document>
    <rank>1</rank>
    <title>
      <![CDATA[ CT Image Management and Communication Services ]]>
    </title>
    <authors>
      <![CDATA[
        Seikaew, K.R.; Jirandorn, N.; Juntaranon, N.; Ariyadech, S.; Sa-Ing, V.; Pattara-Atikom, W.; Thongvigitmanee, S.S.
      ]]>
    </authors>
    <affiliations>
      <![CDATA[
        Dept. of Comput. Eng., Khon Kaen Univ., Khon Kaen, Thailand
      ]]>
    </affiliations>
    <controlledterms>
      <term>
        <![CDATA[ Internet ]]>
      </term>
      <term>
        <![CDATA[ computerised tomography ]]>
      </term>
      <term>
        <![CDATA[ dentistry ]]>
      </term>
      <term>
        <![CDATA[ medical image processing ]]>
      </term>
      <term>
        <![CDATA[ mobile computing ]]>
      </term>
    </controlledterms>
    <thesaurusterm>
      <![CDATA[ SRII Global Conference (SRII), 2012 Annual ]]>
    </thesaurusterm>
    <pubtype>
      <![CDATA[ Conference Publications ]]>
    </pubtype>
    <publisher>
      <![CDATA[ IEEE ]]>
    </publisher>
    <year>
      <![CDATA[ 2012 ]]>
    </year>
    <issn>
      <![CDATA[ 978-1-4673-2318-5 ]]>
    </issn>
  </document>

```

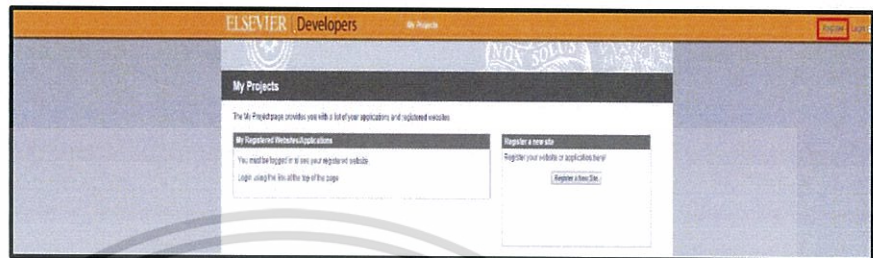
รูปที่ 3.5 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ IEEE Xplore

3.2.2 Scopus

Scopus ฐานข้อมูลบรรณานุกรมและสาระสังเขปครอบคลุมสาขาวิชาต่างๆ ดังนี้ เคมี ฟิสิกส์ คณิตศาสตร์ วิศวกรรมศาสตร์ สังคมวิทยา จิตวิทยา เศรษฐศาสตร์ วิทยาศาสตร์ การแพทย์ ชีวภาพ เกษตรศาสตร์ และสิ่งแวดล้อม สำหรับข้อมูลด้านการวิจัยได้รวบรวมข้อมูลสาขาต่างๆ จากวารสารวิชาการชั้นนำ โดยการดึงข้อมูลผ่าน API ของเว็บไซต์ Scopus ต้องทำการดึงข้อมูลทั้งหมดสองครั้งด้วยกัน คือ การดึงครั้งแรกจะเป็นการดึงข้อมูลผ่าน API โดยการค้นจากชื่อของนักวิจัยจะได้ลิงก์ผลงานของนักวิจัยและข้อมูลเบื้องต้นของผลงาน จากนั้นนำลิงก์ผลงานได้มาไปดึงข้อมูลผลงานอีกครั้งหนึ่ง ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ Scopus มีดังนี้

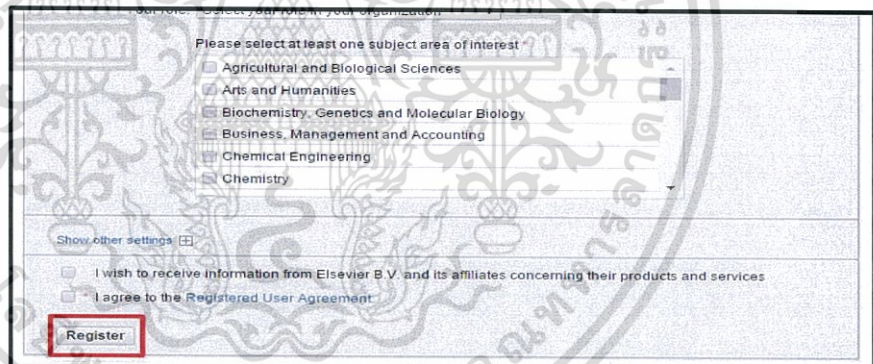
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) การดึงข้อมูลจากเว็บไซต์ “Scopus” จะต้องทำการสมัครสมาชิกเพื่อขอ API key โดยใส่ URL: <http://www.developers.elsevier.com/action/devprojects> ที่เว็บเบราว์เซอร์จะปรากฏหน้าเว็บไซต์ [11] ตามรูปที่ 3.6



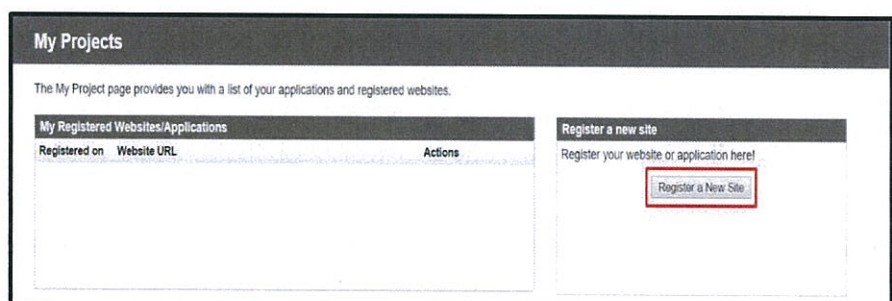
รูปที่ 3.6 หน้าเว็บไซต์ “ELSEVIER”

- 2) ทำการสมัครสมาชิกโดยเลือกปุ่ม “Register” อยู่ด้านมุมขวาบนของเว็บไซต์จะปรากฏหน้าสมัครสมาชิกให้กรอกข้อมูลให้ครบถ้วน แล้วเลือกปุ่ม “Register” ด้านล่างของเว็บไซต์ ตามรูปที่ 3.7



รูปที่ 3.7 หน้าสมัครสมาชิกของ เว็บไซต์ “ELSEVIER”

- 3) ทำการสมัครเพื่อขอ “API Key” โดยเลือกที่ “Register a New Site” ตามรูปที่ 3.8



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ 3.8 การสมัครเพื่อขอ “API Key” ของเว็บไซต์ Scopusด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 4) ใส่ <http://www.scopus.com> ที่ช่อง Website URL แล้วเลือก “Register Site” ตามรูปที่ 3.9

After entering the information below, click Register site. (* = required field)

Website URL:

API Service Agreement

[Click here to read the API Service Agreement.](#)

I agree with the API Service Agreement.

To register this site, you must read and agree with the API Service Agreement.

รูปที่ 3.9 การสร้าง API Key ของเว็บไซต์ Scopus

- 5) เลือกที่ “View API Key” ตามรูปที่ 3.10 จากนั้นจะปรากฏ “API Key” ขึ้นมาตามรูปที่ 3.11

Registered on	Website URL	Actions
15 Nov 2015	http://www.scopus.com	View API Key Edit Delete

รูปที่ 3.10 วิธีการดู “API Key” ที่ใช้ในการดึงข้อมูลของเว็บไซต์ Scopus

API Key

รูปที่ 3.11 “API Key” ที่ใช้ในการดึงข้อมูลของเว็บไซต์ Scopus

- 6) จากนั้นนำ API key ที่ได้มารวมกับ URL โดยตัวอย่างจะทำการค้นหาโดยใช้ชื่อนักวิจัย “Wasan Pattara-atikom” ตามรูปที่ 3.12

```
URL = http://api.elsevier.com/content/search/scopus?apikey=05b638a38993c35196a5afacab2be155&
httpAccept=application/json&query=AUTH%28Wasan+Pattara-atikom%29&r=46758
```

รูปที่ 3.12 ตัวอย่าง URL ที่ใช้ในการดึงข้อมูลผ่าน API ของเว็บไซต์ Scopus

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 7) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบของ JSON ตามรูปที่ 3.13

```
prism:url: "http://api.elsevier.com/content/abstract/scopus_id/84928911022"
dc:identifier: "SCOPUS_ID:84928911022",
eid: "2-s2.0-84928911022",
dc:title: "Road traffic question answering system using ontology",
dc:creator: "Wanichayapong N.",
```

รูปที่ 3.13 ข้อมูลผลงานเบื้องต้นของเว็บไซต์ Scopus

- 8) นำลิงก์ผลงานของนักวิจัย (ข้อมูลใน prism:url:) ของรูปที่ 3.13 ที่ดึงมาได้มาทำการดึงข้อมูลผลงานต่อโดยนำมาใส่ URL รูปที่ 3.14

```
URL = http://api.elsevier.com/content/abstract/scopus_id/84928911022?apiKey=05b638a38993c35196a5afacab2be155&httpAccept=application/json&view=FULL
```

รูปที่ 3.14 ตัวอย่าง URL ที่ใช้ในการดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ Scopus

- 9) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบของ JSON ตามรูปที่ 3.15 โดยผลลัพธ์ข้อมูลที่ทำการดึง ได้แก่ title, author, cited, publish year, keywords

```
coredata: {
  prism:url: "http://api.elsevier.com/content/abstract/scopus_id/84928911022"
  dc:identifier: "SCOPUS_ID:84928911022",
  eid: "2-s2.0-84928911022",
  prism:doi: "10.1007/978-3-318-15615-6_32",
  dc:title: "Road traffic question answering system using ontology",
  prism:aggregationType: "Book Series",
  srxtype: "A",
  citedby-count: "0",
  prism:publicationName: "Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)",
  dc:publisher: "Springer-Verlag service.springer.de",
  source-id: "35674",
  prism:isbn: "9783317156149",
  prism:issn: "16113349",
  prism:volume: "8543",
```

รูปที่ 3.15 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Scopus

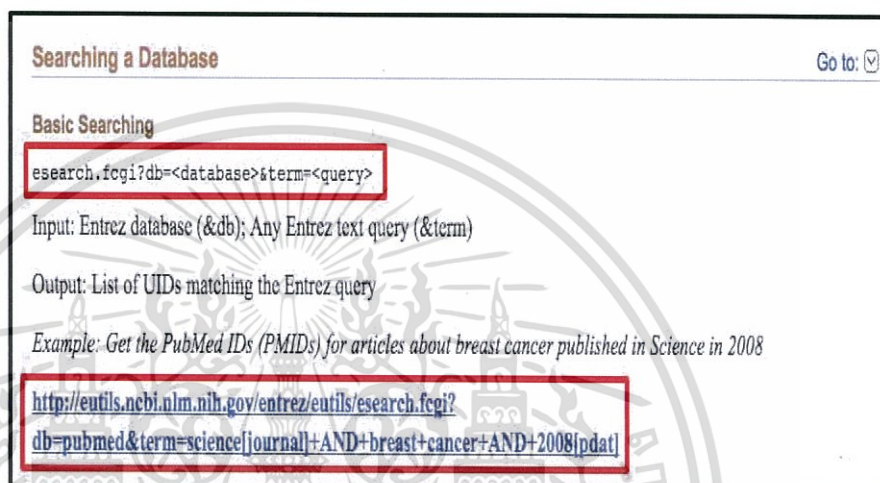
3.2.3 PubMed

PubMed เป็นเครื่องมือค้นหาฐานข้อมูลตัวหนึ่งซึ่งเรียกใช้ข้อมูล จากฐานข้อมูลของ MEDLINE (MEDLINE เป็นฐานข้อมูลที่ครอบคลุมวารสารสาขาชีววิทยาการแพทย์ ที่สำคัญไว้เกือบทุกฉบับ จัดทำโดย U.S. National Library of Medicine (NLM)) ซึ่งประกอบด้วยแหล่งอ้างอิง บทความย่อ และบทความตัวเต็มทางวิทยาศาสตร์และชีวเวชศาสตร์ โดยการดึงข้อมูลผ่าน API ของเว็บไซต์ PubMed นั้นต้องทำการดึงข้อมูลทั้งหมดสองครั้งคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การดึงครั้งแรกจะเป็นการดึงข้อมูลผ่าน API โดยการค้นจากชื่อของนักวิจัย จะได้รายการหมายเลขผลงานของนักวิจัยนั้นๆ จากนั้นนำไอดีผลงานได้มาไปดึงข้อมูลผลงานอีกทีหนึ่ง ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ PubMed มีดังนี้

- 1) ใส่ URL: <http://www.ncbi.nlm.nih.gov/books/NBK25500> ที่เว็บไซต์เบราว์เซอร์ จะปรากฏหน้าวิธีการ และตัวอย่างการดึงข้อมูลผ่าน API ของเว็บไซต์ PubMed [13] ตามรูปที่ 3.16



รูปที่ 3.16 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ PubMed

- 2) จากนั้นทำการแทน ชื่อ-นามสกุล ของนักวิจัยลงใน URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ PubMed ตามรูปที่ 3.17 โดยใช้ชื่อ “Alissara Reungsang” เป็นตัวอย่างในการค้นหาข้อมูล

URL = [http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=Reungsang,Alissara\[author\]&retmode=json](http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=Reungsang,Alissara[author]&retmode=json)

รูปที่ 3.17 ตัวอย่าง URL ที่ใช้ดึงข้อมูลไอดีผลงานผ่าน API ของเว็บไซต์ PubMed

- 3) ข้อมูลที่ดึงได้จะเป็นรายการหมายเลขผลงานของนักวิจัยในรูปแบบของ JSON ตามรูปที่ 3.18

```

- header: {
  type: "esearch",
  version: "0.3"
},
- esearchresult: {
  count: "14",
  retmax: "14",
  retstart: "0",
  - idlist: [
    "25079849",
    "24412153",
    "23845708",
    "23530362",
    "23500404",
    "22286712",
    "22173656",
    "21793396",
    "21549227",
    "21502737",
    "20107983",
    "19092209",
    "19093464",
    "19086538"
  ],
  - translationset: [
    - {
      from: "Reungsang, Alissara[author]",
      to: "Reungsang, Alissara[Full Author Name]"
    }
  ],
}

```

รูปที่ 3.18 ข้อมูลรายการโอตีมผลงานของนักวิจัยในรูปแบบของ JSON ของเว็บไซต์ PubMed

- 4) จากนั้นนำหมายเลขผลงานของนักวิจัยที่ดึงได้ มาทำการดึงข้อมูลผ่าน API โดยนำหมายเลขผลงานของนักวิจัยแทนลงใน URL โดยจะใช้หมายเลขผลงาน "25079849" เป็นตัวอย่างในการดึงข้อมูล ตามรูปที่ 3.19

URL = <http://www.ncbi.nlm.nih.gov/pubmed/?term=25079849&report=xml>

รูปที่ 3.19 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ PubMed

- 5) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบของ XML ตามรูปที่ 3.20 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ title, author, publish year, keywords

```

<PubmedArticle>
  <MedlineCitation Owner="NLM" Status="MEDLINE">
    <PMID Version="1">25079849</PMID>
    <DateCreated>
      <Year>2014</Year>
      <Month>08</Month>
      <Day>01</Day>
    </DateCreated>
    <DateCompleted>
      <Year>2014</Year>
      <Month>11</Month>
      <Day>10</Day>
    </DateCompleted>
    <Article PubModel="Print">
      <Journal>
        <ISSN IssnType="Print">1001-0742</ISSN>
        <JournalIssue CitedMedium="Print">
          <Volume>26</Volume>
          <Issue>6</Issue>
          <PubDate>
            <Year>2014</Year>
            <Month>Jun</Month>
            <Day>1</Day>
          </PubDate>
        </JournalIssue>
        <Title>Journal of environmental sciences (China)</Title>
        <ISOAbbreviation>J Environ Sci (China)</ISOAbbreviation>
      </Journal>

```

รูปที่ 3.20 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ PubMed

- 6) ทำการดึงข้อมูล “Cited” ของผลงาน โดยจะใช้ไอดีผลงาน “21502737” เป็นตัวอย่างในการดึงข้อมูล ตามรูปที่ 3.21

```

http://www.ncbi.nlm.nih.gov/pubmed/?pSl=AjaxServer&portlets=id=pmfavad:id=pmdiscora:db=pubmed:
pmid=21502737:id=pmcommonsad:page=none:id=pmdiscopmc:db=pubmed:pmid=21502737:id=pmdisc
ostruct:db=pubmed:pmid=21502737:id=pmdiscodblinks:db=pubmed:pmid=21502737:id=recentactivity&
load=yes&page=full

```

รูปที่ 3.21 ตัวอย่าง URL ที่ใช้ดึงข้อมูล “Cited” ของผลงานของเว็บไซต์ PubMed

- 7) ทำการดึงข้อมูล Source Code ของหน้าเว็บไซต์โดยใช้ URL ตามรูปที่ 3.21

```

seeall&logdbfrom=pubmed&ncbi_uid=21502737">See reviews...</a><a
omed_pubmed&from_uid=21502737"
s seeall&logdbfrom=pubmed&ncbi_uid=21502737">See all...</a></div></div>
<div class="portlet_title"><h3><span>Cited by 6 PubMed Central articles</span>
class="portlet_shutter" title="Show/hide content" remembercollapsed="true"
</div><div class="portlet_content"><ul><li class="brieflinkpopper two_line"><a
/26300864"
citedinpmarticles&logdbfrom=pubmed&ncbi_uid=21502737&link_uid=263
es in total_ and cultururable-bacterial assemblages in South African aquatic

```

รูปที่ 3.22 ตัวอย่าง source code ของหน้าเว็บไซต์ PubMed

จากรูปที่ 3.22 ทำการดึงข้อมูลตัวเลขตรงส่วนของ Cited By ออกมาโดยการใช้คำสั่ง substr() ดึงข้อมูลออกมาบางส่วน แล้วใช้ฟังก์ชัน preg_match() เอกสารนี้เป็นเอกสารที่สงวนไว้ในการดึงแค่ตัวเลขออกมาซึ่งข้อมูลที่ดึงจะได้เลข 6 ออกมาซึ่งประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.4 Slide Share

SlideShare เป็น website ที่ให้บริการฝากไฟล์ ประเภท งานนำเสนอและวิดีโอ สามารถนำไปแชร์ให้ผู้อื่นได้เข้ามาดู หรือ Download ไปใช้ได้นอกจากนี้ยังสามารถค้นหาไฟล์เอกสาร (word) หรือ งานนำเสนอ (Powerpoint) ที่ผู้อื่นสร้างไว้มากมายมาประกอบกิจกรรมการเรียนการสอนได้ โดยการดึงข้อมูลผ่าน API ของเว็บไซต์ Slideshare จะต้องทำการสมัครสมาชิกเพื่อร้องขอ“คีย์” โดยพารามิเตอร์ ที่ใช้ในการดึงข้อมูลผ่าน API ต้องใช้พารามิเตอร์ทั้งหมด 3 ตัว [14] ตามตารางที่ 3.1

ตารางที่ 3.1 แสดงพารามิเตอร์ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ SlideShare

พารามิเตอร์	รายละเอียด	หมายเหตุ
api_key	API Key ที่ SlideShare ทำการสร้างให้	ได้จากการร้องขอ “คีย์” จากทางเว็บไซต์
ts	เวลาปัจจุบัน ในรูปแบบ “Unix TimeStamp”	ต้องเขียนโค้ดเรียกใช้งานฟังก์ชัน time()
hash	SHA1 การเอา shared secret ต่อกับ timestamp (ts) ลำดับเงื่อนไขการเรียงต่อกันเป็นสิ่งสำคัญ ตัวอย่าง SHA1 (sharedsecret + timestamp)	“shared secret” ได้จากการร้องขอ“คีย์” จากทางเว็บไซต์

สำหรับขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ SlideShare มีดังนี้

- 1) การดึงข้อมูลจากเว็บไซต์ SlideShare จะต้องทำการสมัครสมาชิกเพื่อขอ API key โดยใส่ URL: <http://www.slideshare.net/> ที่เว็บเบราว์เซอร์ แล้วคลิกที่คำว่า “Signup” ตรงมุมบนซ้ายของหน้าเว็บไซต์ตามรูปที่ 3.23



รูปที่ 3.23 หน้าเว็บไซต์ SlideShare เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อวัตถุประสงค์เฉพาะของผู้ใช้เท่านั้น ไม่สามารถเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารได้ นอกจากนี้ เอกสารนี้ยังเป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อวัตถุประสงค์เฉพาะของผู้ใช้เท่านั้น ไม่สามารถเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารได้ นอกจากนี้ เอกสารนี้ยังเป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อวัตถุประสงค์เฉพาะของผู้ใช้เท่านั้น ไม่สามารถเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารได้

2) กรอกข้อมูลให้ครบถ้วน แล้วคลิกปุ่ม “Signup” ตามรูปที่ 3.24

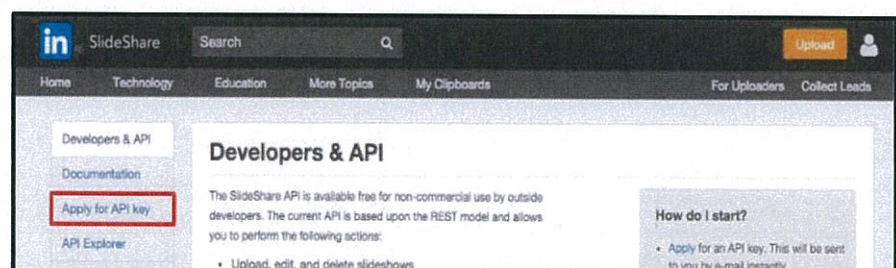
รูปที่ 3.24 การสมัครสมาชิกของเว็บไซต์ SlideShare

3) หลังจาก Signup เสร็จแล้ว จะกลับมาหน้าหลักของเว็บไซต์ให้คลิกที่ “Dev & API” ที่อยู่ส่วนท้ายของเว็บไซต์ ตามรูปที่ 3.25 จะปรากฏหน้าเว็บไซต์ แสดงรายละเอียด การดึงข้อมูลผ่าน API ของเว็บไซต์ SlideShare



รูปที่ 3.25 การเข้าถึงหน้าเว็บไซต์ที่แสดงรายละเอียดการดึงข้อมูลผ่าน API ของเว็บไซต์ SlideShare

4) ทำการขอ API Key โดยคลิกที่ปุ่ม “Apply to API key” ตามรูปที่ 3.26



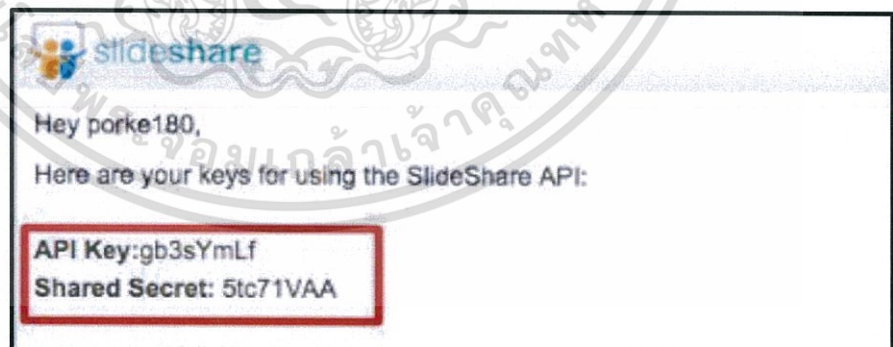
รูปที่ 3.26 วิธีการเข้าถึงหน้าเว็บไซต์ที่ขอ API Key ของเว็บไซต์ SlideShare

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) จะปรากฏหน้าเว็บไซต์ให้กรอกคำขอ API Key

รูปที่ 3.27 หน้ากรอกคำขอ API Key ของเว็บไซต์ SlideShare

กรอกข้อมูลให้ครบถ้วน จากนั้นคลิกปุ่ม “Apply for API Key” ตามรูปที่ 3.27 API key คือ “gb3sYmLf” และ Shared secret คือ “5tc71VAA” จะถูกส่งไปที่ email ที่ใช้สมัครสมาชิก ตามรูปที่ 3.28



รูปที่ 3.28 “API key” และ “Shared secret” ของเว็บไซต์ SlideShare

6) จากนั้นนำ API key, Shared Secret และ st ที่ได้มารวมกับ URL ที่ใช้ดึงข้อมูลผลงานของนักวิจัยผ่าน API โดยจะใช้ชื่อ “Wasan Pattara-atikom” เป็นตัวอย่าง จะได้ URL ตามรูปที่ 3.29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้拿去ใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
https://www.slideshare.net/api/2/search_slideshows?q=Wasan%20Pattara-atikom&page=1&items_per_page=16&lang=**&sort=relevance&upload_date=any&fileformat=all&file_type=all&cc=1&cc_adapt=1&cc_commercial=1&api_key=gb3sYmLf&hash=f24d00cdd42b450c31b14df70dc8f156b2d50a34&ts=1446538107
```

รูปที่ 3.29 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ SlideShare

- 7) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบ XML ตามรูปที่ 3.30 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ Title, Description, URL, Created, Updated, DownloadUrl

```
This XML file does not appear to have any style information associated with it. The document tree is shown below.
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<Slideshow>
  <Meta>...</Meta>
  <SlideShow>
    <ID>43009632</ID>
    <Title>Irpc Process license-PTT forum</Title>
    <Description>
      IRPC Process license presentation slide was presented to PTT forum 2012
    </Description>
    <Status>2</Status>
    <Username>jakkawal</Username>
    <URL>
      http://www.slideshare.net/jakkawal/irpc-process-licenseptt-forum
    </URL>
    <ThumbnailURL>...</ThumbnailURL>
    <ThumbnailSize>[170,130]</ThumbnailSize>
    <ThumbnailSmallURL>...</ThumbnailSmallURL>
    <ThumbnailXLargeURL>...</ThumbnailXLargeURL>
    <Embed>...</Embed>
    <Created>2014-12-25 21:31:04 UTC</Created>
    <Updated>2014-12-25 21:32:32 UTC</Updated>
    <Language>eng</Language>
    <Format>pdf</Format>
    <Download>1</Download>
    <DownloadURL>
      http://s3.amazonaws.com/ppt-download/irpc-processlicense-pttforumrev-141225213104-conversion-gate01.pdf?response-content-disposition=attachment&signature=80cfivlfcarc0MSAGURfzSUNK0w3JdeExpires=1446538429&AWSAccessKeyId=AKIAIA7QTBORZLDUZR1Q
    </DownloadURL>
    <SecretKey>angsfkvwB05m</SecretKey>
    <SlideshowEmbedUrl>...</SlideshowEmbedUrl>
    <SlideshowType>0</SlideshowType>
    <InContest>0</InContest>
  </SlideShow>
</Slideshow>
```

รูปที่ 3.30 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ SlideShare

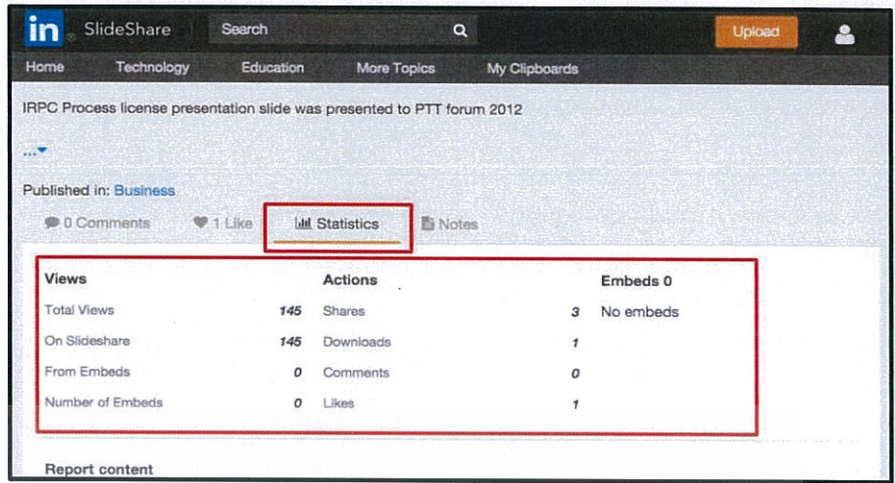
- 8) การดึงข้อมูลผ่าน API เพียงอย่างเดียวไม่สามารถได้ข้อมูลที่ต้องการได้ครบถ้วน จึงนำข้อมูลในส่วนของ URL ที่สามารถดึงมาได้ของแต่ละผลงานมาทำการดึงข้อมูลต่ออีกครั้งหนึ่ง โดยจะใช้ตัวอย่าง URL ตามรูปที่ 3.31

```
URL = http://www.slideshare.net/jakkawal/irpc-process-licenseptt-forum
```

รูปที่ 3.31 ตัวอย่าง URL ที่ใช้ในการดึงข้อมูลของเว็บไซต์ SlideShare

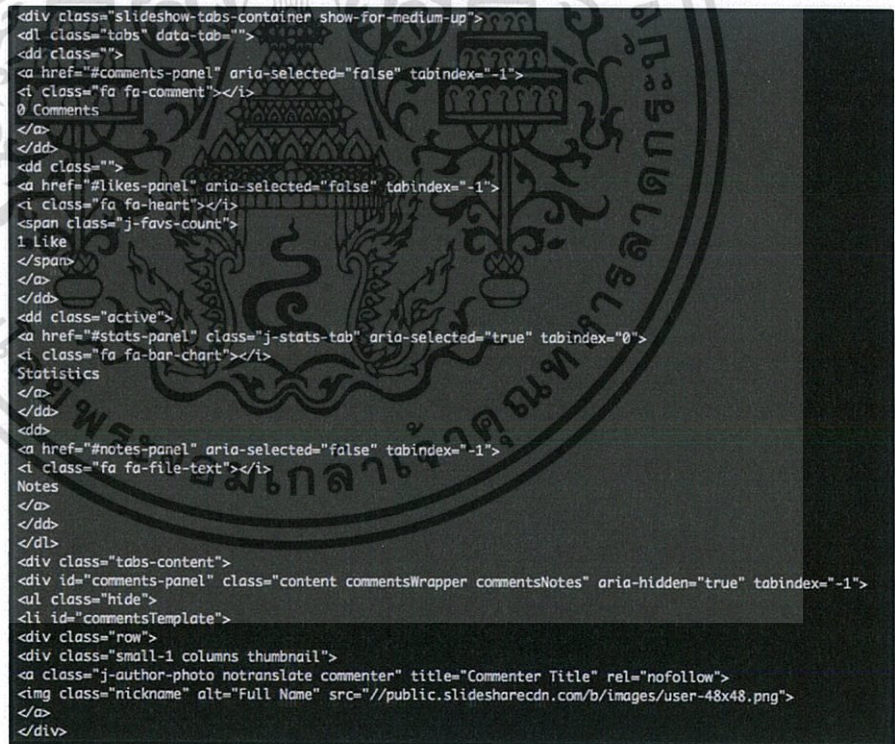
- 9) ทำการดึงข้อมูล Source Code หน้าเว็บไซต์ข้อมูลผลงานจากข้อมูล URL มาด้วยฟังก์ชัน file_get_html() โดยส่วนที่เราต้องการจะอยู่ในส่วนของ Statistics ตามรูปที่ 3.32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.32 หน้าเว็บไซต์ผลงานในส่วนของข้อมูล Statistics ของเว็บไซต์ SlideShare

10) ข้อมูลที่ดึงได้ก็จะอยู่ในรูปแบบของ XML ตามรูปที่ 3.33 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ Views, Shares, Downloads, Comments, Likes



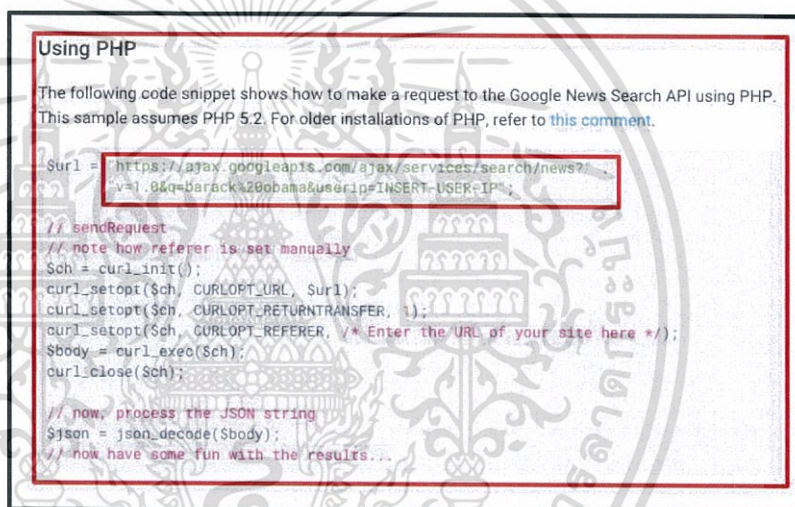
รูปที่ 3.33 ข้อมูล Statistics ของผลงานในรูปแบบของ XML ของเว็บไซต์ SlideShare

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.5 Google News Search

Google News Search เป็นเครื่องมือที่ให้บริการค้นหาข้อมูลบนอินเทอร์เน็ต (Search Engine) ของเว็บไซต์ Google ในหมวดของกลุ่มข่าว เป็นการค้นหาข่าวสารจากกลุ่มสมาชิกที่ใช้บริการ Google News Groups เพื่อรับส่งข่าวสารกันเองระหว่างสมาชิก โดยมี การระบุชื่อผู้เขียน หัวข้อข่าว วันที่และเวลาที่โพสต์ข่าว ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search มีดังนี้

- 1) ใส่ URL: “https://developers.google.com/news-search/v1/jsondevguide#using-php” ที่เว็บเบราว์เซอร์ จะปรากฏหน้าวิธีการ และ ตัวอย่างการดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search [15] ตาม รูปที่ 3.34



```

Using PHP
The following code snippet shows how to make a request to the Google News Search API using PHP.
This sample assumes PHP 5.2. For older installations of PHP, refer to this comment.

$url = "https://ajax.googleapis.com/ajax/services/search/news?v=1.0&q=barack%20obama&userip=INSERT-USER-IP";

// sendRequest
// note how referer is set manually
$ch = curl_init();
curl_setopt($ch, CURLOPT_URL, $url);
curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
curl_setopt($ch, CURLOPT_REFERER, /* Enter the URL of your site here */);
$body = curl_exec($ch);
curl_close($ch);

// now, process the JSON string
$json = json_decode($body);
// now have some fun with the results...

```

รูปที่ 3.34 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search

จากรูปที่ 3.34 เราจะได้ URL ที่ใช้ในการดึงข้อมูลผ่าน API ที่ใช้ชื่อในการ ค้นหาโดยจากตัวอย่างนั้นทำการค้นหาโดยใช้ชื่อ “barack obama”

- 2) จากนั้นทำการแทน ชื่อ-นามสกุล ของนักวิจัยลงใน URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search ตามรูปที่ 3.35 โดยใช้ชื่อ “Wasan Pattara-atikom” เป็นตัวอย่างในการค้นหาข้อมูล

URL = https://ajax.googleapis.com/ajax/services/search/news?v=1.0&q=%22Wasan+ Pattara-atikom%22&userip

รูปที่ 3.35 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google News

เอกสารนี้เป็นเอกสารที่สงวนไว้ Searchการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 3) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบ JSON ตามรูปที่ 3.36 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ content, unescapedUrl, titleNoFormatting, publisher, language, publishedDate

```

{
  - responseData: {
    - results: [
      - {
        GsearchResultClass: "GnewsSearch",
        clusterUrl: "",
        content: "... อินเทอร์เน็ต หรือ ทราฟฟ์ แอปพลิเคชัน (Traffic Application Programming Interface :Traffy API)",
        unescapedUrl: "http://www.thairath.co.th/content/175084",
        url: "http%3A%2F%2Fwww.thairath.co.th%2Fcontent%2F175084",
        title: "เน็ตเทค สมชายภักTraffy API เปิดทางเอกชนจ่ายหมิ่นบาทต่อบลูโง้งงาน",
        titleNoFormatting: "เน็ตเทค สมชายภักTraffy API เปิดทางเอกชนจ่ายหมิ่นบาทต่อบลูโง้งงาน",
        location: "",
        publisher: "ไทยรัฐ",
        publishedDate: "Sun, 29 May 2011 19:01:07 -0700",
        signedRedirectUrl: "http://news.google.com/news/url?sa=T&t2=th&fd=S&url=http://www.thairath",
        language: "th",
        - image: {
          url: "http://www.thairath.co.th/media/NjpUs24nCOkx5e1GFLM4ZaNoTauapivdWF7x3f01VJ9.jpg",
          tbUrl: "http://t3.gstatic.com/images?q=tbn:AND9GcRzSIMKRfcsY1DUa7pmR_c336UgPyS5WCT4coO11",
          originalContextUrl: "http://www.thairath.co.th/content/175084",
          publisher: "ไทยรัฐ",
          cbWidth: 80,
          cbHeight: 48
        }
      }
    ]
  }
}

```

รูปที่ 3.36 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Google News Search

3.2.6 Google Video Search

Google Video Search เป็นเครื่องมือที่ให้บริการค้นหาข้อมูลบนอินเทอร์เน็ต (Search Engine) ของเว็บไซต์ Google ในหมวดของวิดีโอ เป็นการค้นหาข้อมูลเกี่ยวกับวิดีโอจากกลุ่มสมาชิกที่ใช้บริการ Google Video โดยมีการระบุชื่อผู้อัปโหลดวิดีโอ หัวข้อวิดีโอ วันที่และเวลาที่โพสต์วิดีโอ ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ Google Video Search มีดังนี้

- 1) ใส่ URL : “https://developers.google.com/video-search/v1/Jsondevguide#using-php” ที่เว็บเบราว์เซอร์ จะปรากฏหน้าวิธีการ และ ตัวอย่างการดึงข้อมูลผ่าน API ของเว็บไซต์ Google Video Search [16] ตาม รูปที่ 3.37

```

Using PHP
The following code snippet shows how to make a request to the Google Video Search API using PHP.
This sample assumes PHP 5.2. For older installations of PHP, refer to this comment.

$url = "https://ajax.googleapis.com/ajax/services/search/video?v=1.0&q=barack%20obama&userip=INSERT-USER-IP";

// sendRequest
// note how referer is set manually
$ch = curl_init();
curl_setopt($ch, CURLOPT_URL, $url);

```

รูปที่ 3.37 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ Google Video Search

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 2) จากนั้นทำการแทน ชื่อ-นามสกุล ของนักวิจัยลงใน URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google Video Search ตามรูปที่ 3.38 โดยใช้ชื่อ “Kasemsak Uthaichana” เป็นตัวอย่างในการค้นหาข้อมูล

URL: <https://ajax.googleapis.com/ajax/services/search/video?v=1.0&q=%22Kasemsak+Uthaichana%22&userip>

รูปที่ 3.38 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google News Search

- 3) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบ JSON ตามรูปที่ 3.39 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ titleNoFormatting, content, publisher, published, tbUrl, tbWidth, tbHeight, videoType, url, playUrl, duration

```

- results: [
  {
    GsearchResultClass: "GvideoSearch",
    title: "นักทำขนมเค้กไทย (1) - YouTube",
    titleNoFormatting: "นักทำขนมเค้กไทย (1) - YouTube",
    published: "Sat, 04 Apr 2015 09:04:20 -0700",
    content: "Chiang Mai Maker Party. Free style talk session. Topic : นักทำขนมเค้กไทย Speaker :",
    publisher: "youtube.com",
    tbUrl: "https://img.youtube.com/vi/JXam7Rz6xXE/default.jpg?h=90&w=120&sig=_E4z6c1MGvh4eB5Xvrc",
    tbWidth: "120",
    tbHeight: "90"
  }
]

```

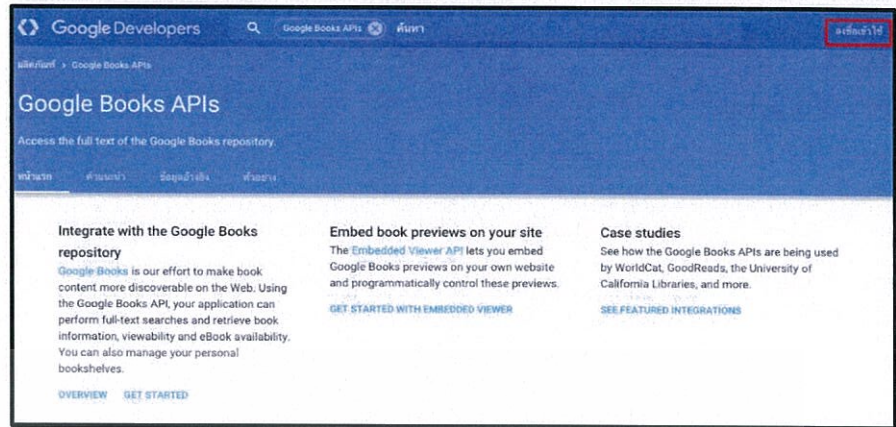
รูปที่ 3.39 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Google Video Search

3.2.7 Google Books APIs

Google Books APIs เป็นเครื่องมือที่ให้บริการค้นหาและสืบค้นข้อมูลบนอินเทอร์เน็ต (Search Engine) ของเว็บไซต์ Google ในหมวดของหนังสือ เป็นการค้นหาข้อมูลเกี่ยวกับหนังสือจากบริการ Google Books โดยมีการระบุชื่อผู้แต่ง หัวข้อของหนังสือ รายละเอียดเกี่ยวกับหนังสือ ภาษาที่ใช้และวันที่ตีพิมพ์ โดยการดึงข้อมูลผ่าน API ของ “Google Books API” นั้นจะแตกต่างจากการดึงข้อมูล “Google News Search” และ “Google Video Search” เนื่องจากต้องมีการสมัครสมาชิกเพื่อสร้าง API key ด้วย ซึ่งนำมาใช้ในการดึงข้อมูลผ่าน API โดยขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ Google Books APIs มีดังนี้

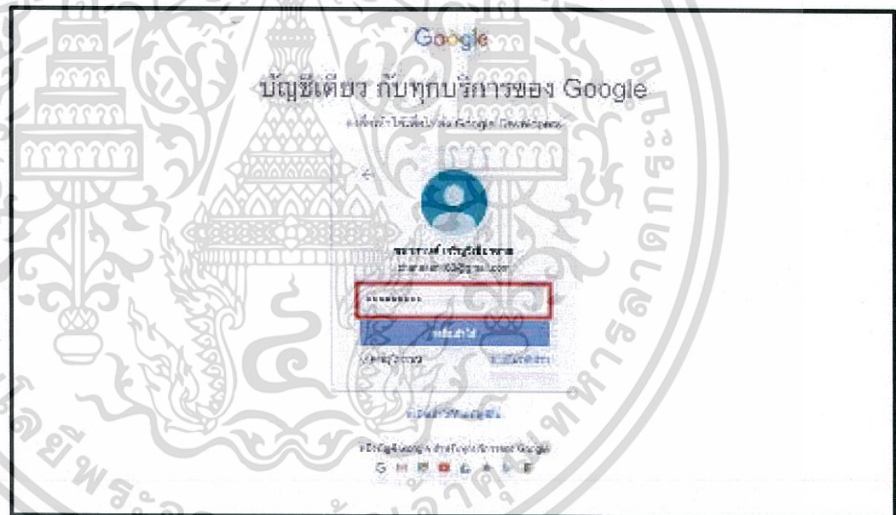
- 1) ใส่ URL: “<https://developers.google.com/books/?hl=th>” ที่เว็บเบราว์เซอร์คลิก “ลงชื่อเข้าใช้” บนมุมขวาของเว็บไซต์เพื่อลงชื่อเข้าใช้งาน [17] ตามรูปที่ 3.40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.40 หน้าเว็บไซต์ Google Books APIs

- 2) กรอก Email Address ที่เชื่อมกับบัญชีกับทาง “Google” แล้วคลิกปุ่ม “ถัดไป” ตามรูปที่ 3.41



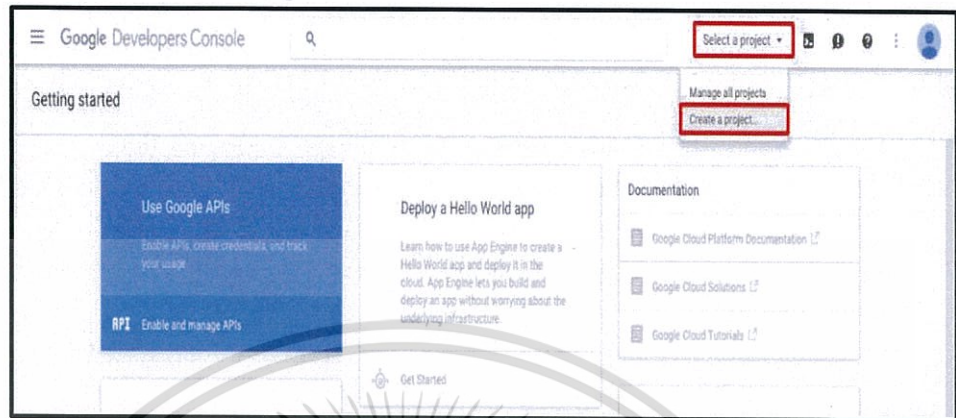
รูปที่ 3.41 หน้าลงชื่อเข้าใช้ของเว็บไซต์ Google Books APIs

- 3) คลิก “Google Developers Console” ตรงส่วนท้ายของเว็บไซต์ ตามรูปที่ 3.42



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ 3.42 วิธีการเข้าหน้า Google Developers Console ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 4) สร้างโปรเจกต์ โดยการเลือกที่ “Select a Project” จากนั้นเลือกที่ “Create a Project” ตามรูปที่ 3.43



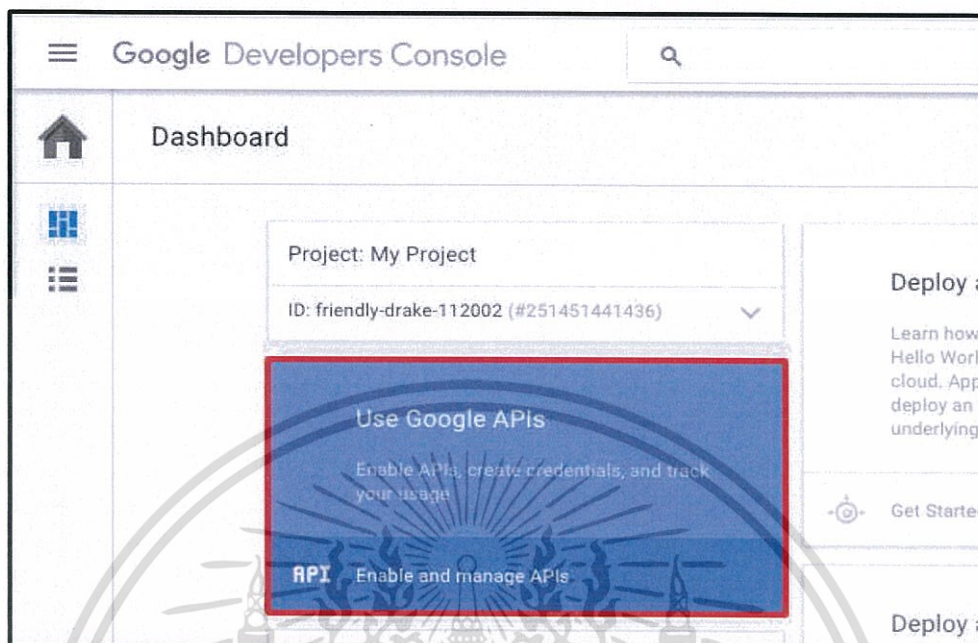
รูปที่ 3.43 วิธีการเข้าหน้าเว็บไซต์ที่ใช้สร้างโปรเจกต์ ของเว็บไซต์ Google Developers Console

- 5) จะมีหน้าต่าง “New Project” ขึ้นมาทำการกรอกข้อมูลให้เรียบร้อย แล้วเลือก “Create” ตามรูปที่ 3.44

รูปที่ 3.44 วิธีการสร้างโปรเจกต์ของเว็บไซต์ Google Developers Console

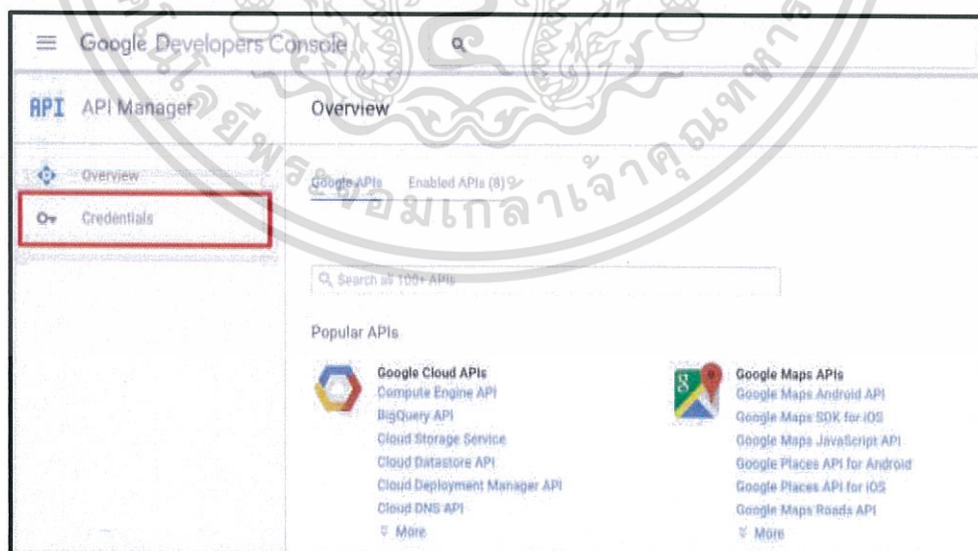
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6) คลิกที่ “Use Google APIs” เพื่อจัดการเกี่ยวกับ API ตามรูปที่ 3.45



รูปที่ 3.45 วิธีการเข้าหน้าเว็บไซต์ที่ใช้จัดการ API ของเว็บไซต์ Google Developers Console

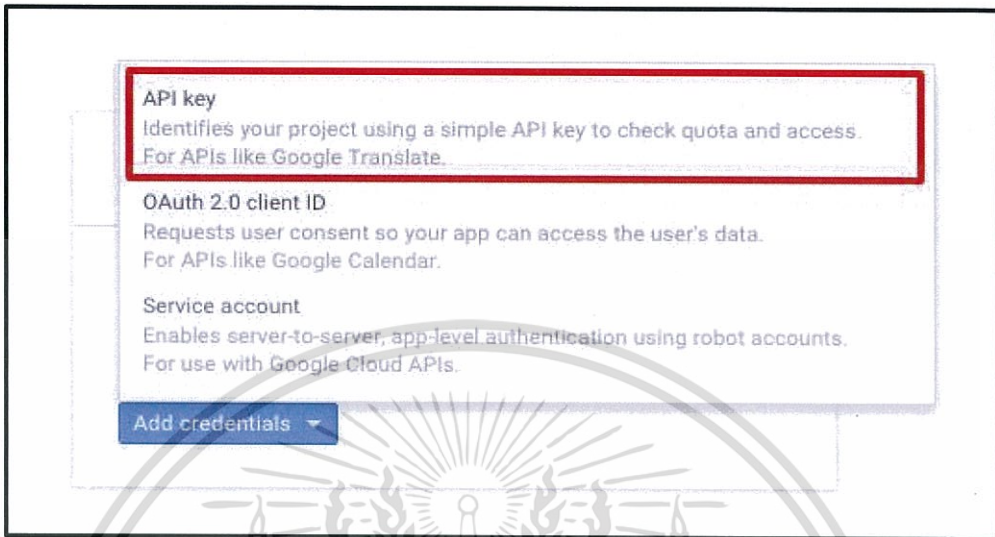
7) การสร้าง API Key เพื่อนำมาใช้ในการสืบค้นข้อมูลผ่าน API ของ Google - คลิก “Credentials” ที่เมนูด้านซ้ายของเว็บไซต์ ตามรูปที่ 3.46



รูปที่ 3.46 การเลือกเมนู “Credentials” ของเว็บไซต์ Google Developers Console

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- คลิกปุ่ม “Add credentials” แล้ว เลือก API Key ตามรูปที่ 3.47



รูปที่ 3.47 การเพิ่ม “Credentials” ของเว็บไซต์ Google Developers Console

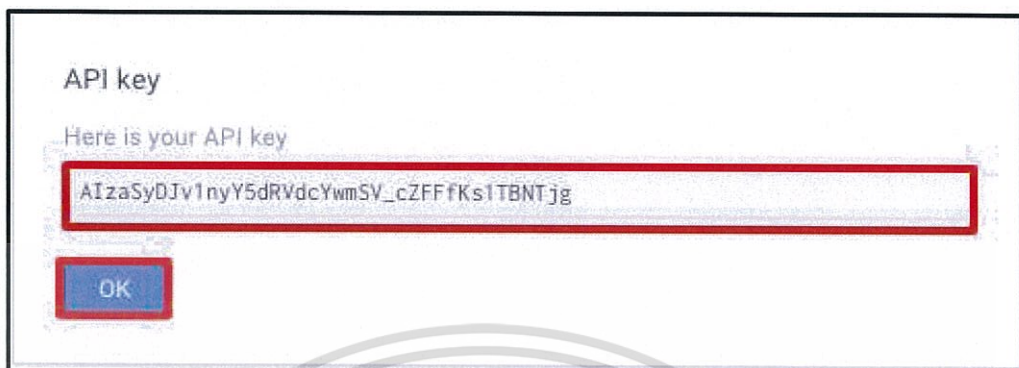
- เลือก “Server key” แล้วตั้งชื่อ API Key ที่ช่อง “Name” จากนั้นคลิกปุ่ม “Create” ตามรูปที่ 3.48



รูปที่ 3.48 การสร้าง API Key ของเว็บไซต์ Google Developers Console

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

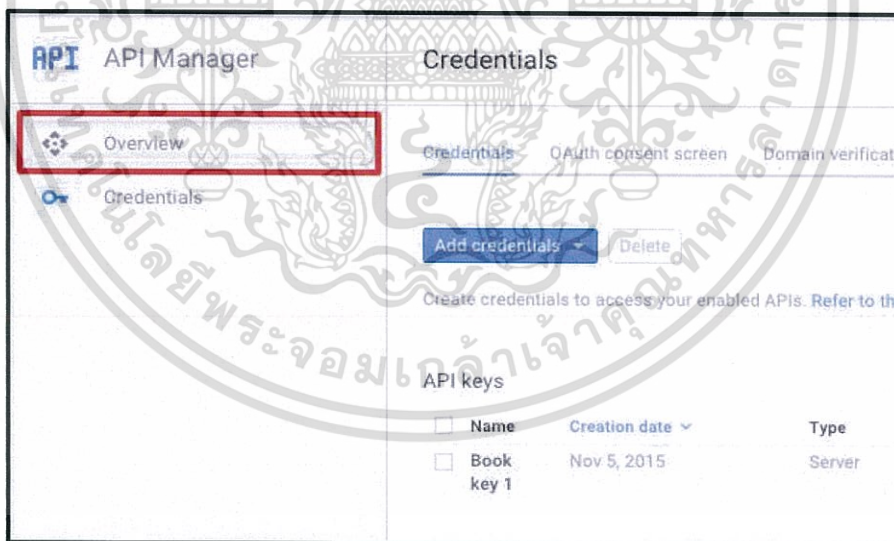
- จะได้ API Key ตามรูปที่ 3.49 เพื่อใช้ในการดึงข้อมูลผ่าน API ของเว็บไซต์ Google Books APIs



รูปที่ 3.49 API Key ของเว็บไซต์ Google Developer Console

- 8) การทำให้ API Key ของโปรเจกต์สามารถใช้งานในส่วนของ Books API ได้นั้นต้องทำการอนุญาตให้ใช้งานก่อน เนื่องจาก Google มีการให้บริการ API ในหลายส่วนจึงต้องมีการขออนุญาตเฉพาะส่วนที่ต้องการใช้งานเท่านั้น

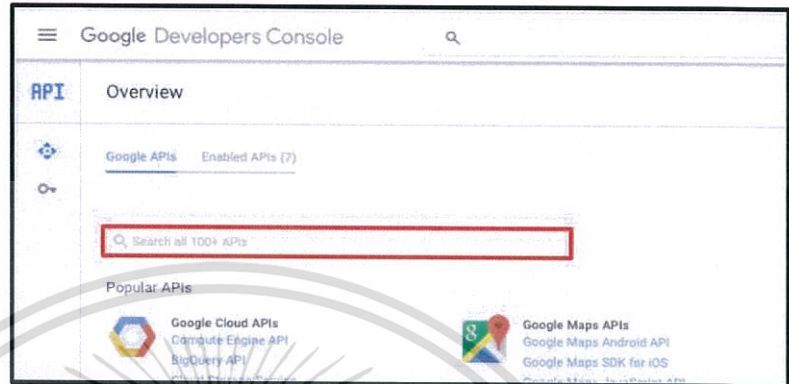
- เลือกเมนู “Overview” ที่เมนูด้านซ้ายของเว็บไซต์ ตามรูปที่ 3.50



รูปที่ 3.50 การเลือกเมนู “Overview” ของเว็บไซต์ Google Developer Console

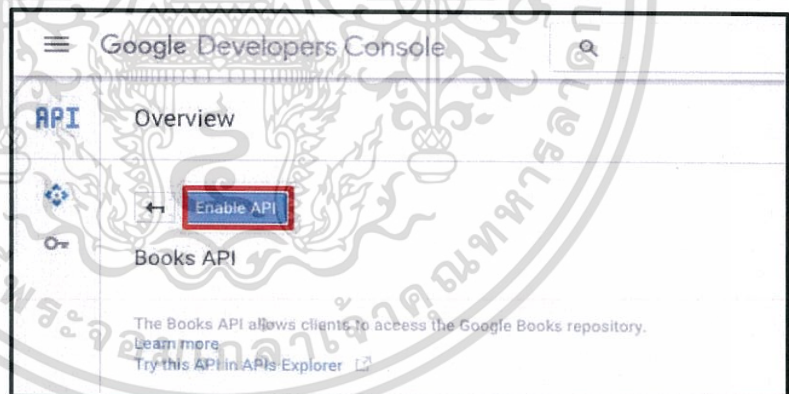
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ใส่คำว่า “Books API” ลงในช่องค้นหา เพื่อค้นหา Books API ตามรูปที่ 3.51



รูปที่ 3.51 การค้นหา Books API ของเว็บไซต์ Google Developer Console

- เลือก “Books API” แล้วคลิกปุ่ม “Enable” เพื่ออนุญาตให้ API Key สามารถใช้ Books API ได้ ตามรูปที่ 3.52



รูปที่ 3.52 การทำให้ API Key ของโปรเจกต์สามารถใช้งาน Books API ได้

- จากนั้นนำ API Key ที่ได้มารวมกับ URL ที่ใช้ดึง API โดยใช้ชื่อ “Nattapong Swangmuang” เป็นตัวอย่างในการค้นหาข้อมูล ตามรูปที่ 3.53

```
URL = https://www.googleapis.com/books/v1/volumes?q=inauthor%3A%22Nattapong+Swangmuang%22&key=AlzaSyDJv1nyY5dRVdcYwmSV_cZFFfKs1TBNTJg&maxResults=40
```

รูปที่ 3.53 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Google Books

APIs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 10) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบ JSON ตามรูปที่ 3.54 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ title, authors, publishedDate, language, previewLink

```

{
  kind: "books#volumes",
  totalItems: 1,
  items: [
    - {
      kind: "books#volume",
      id: "yRk-IcEACAAJ",
      etag: "L52Hr7NPePT",
      selfLink: "https://www.googleapis.com/books/v1/volumes/yRk-IcEACAAJ",
      volumeInfo: {
        title: "A Location Fingerprint Framework Towards Efficient Wireless Indoor Positioning Systems.",
        authors: [
          "Nattapong Srasameang"
        ],
        publishedDate: "2011-09-04",
        description: "Location of mobile computers, potentially indoors, is essential information to enable location-aware applications in wireless pervasive computing. The popularity of wireless local area networks (WLANs) inside and around buildings makes positioning systems based on readily available received signal strength (RSS) from access points (APs) desirable. The fingerprinting technique associates location-dependent characteristics of the received signal strength (RSS) from access points (APs) with the location of the mobile device. This paper presents a location fingerprinting framework for indoor positioning systems. The framework consists of a location fingerprinting algorithm and a location estimation algorithm. The location fingerprinting algorithm is based on the received signal strength (RSS) from access points (APs) and the location of the mobile device. The location estimation algorithm is based on the location fingerprinting algorithm and the location of the mobile device. The location fingerprinting algorithm is based on the received signal strength (RSS) from access points (APs) and the location of the mobile device. The location estimation algorithm is based on the location fingerprinting algorithm and the location of the mobile device."
      }
    }
  ]
}

```

รูปที่ 3.54 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Google Books APIs

3.2.8 Inspire-Hep

Inspire-Hep เครื่องมือในการสืบค้นฐานข้อมูลที่รวบรวมข้อมูลเอกสารทางวิทยาศาสตร์ที่เกี่ยวข้องกับ ฟิสิกส์พลังงานสูง (high energy physics) ของฐานข้อมูล Stanford Physics Information Retrieval System (SPIRES) ซึ่งเป็นฐานข้อมูลหลักสำหรับงานเขียนทางด้านฟิสิกส์พลังงานสูงตั้งแต่ปี 1970 ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ Inspirehep มีดังนี้

- 1) ใส่ URL: <http://inspirehep.net/info/hep/api?ln=en> ที่เว็บเบราว์เซอร์ ปรากฏหน้าวิธีการ และ ตัวอย่างการดึงข้อมูลผ่าน API ของเว็บไซต์ Inspirehep [18] ตามรูปที่ 3.55

Inspire API query format

The API requests have the following general form:

```
GET /search?p=...&of=...&ot=...&jrec=...&rg=...
```

To return information on a single record, simply replace /search with /record/

p = pattern (query)

This is the query in the Inspire search syntax. All search features and operators familiar from the Inspire web interface and documented in the [manual](#) and complex queries are possible.

of = output format

The format of the response sent back to the client. There are two choices, of=xm for (MARC-) XML or of=json for JSON. The XML response fragments thereof when individual fields are selected via the ot parameter

ot = output tags

Select (filter) specific tags from the MARCXML response. This option takes a comma separated list of MARC tags. Valid MARC tags for [MARCXML](#).

รูปที่ 3.55 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ Inspirehep ใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 2) จากนั้นทำการแทน ชื่อ-นามสกุล ของนักวิจัยลงใน URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Inspirehep ตามรูปที่ 3.56 โดยใช้ชื่อ “Burin Gumjudpai” เป็นตัวอย่างในการค้นหาข้อมูล

URL = <http://inspirehep.net/search?p=Burin+Gumjudpai&of=json>

รูปที่ 3.56 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ Inspirehep

- 3) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบของ JSON ตามรูปที่ 3.57 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ recid, Title, author, publication_info, keywords, isbn, cited

```

keywords: [
  {
    institute: "author",
    term: "Power-law cosmology"
  },
  {
    institute: "author",
    term: "quintessence"
  }
],
publication info: {
  volume: "A28",
  pagination: "1350122",
  title: "Mod.Phys.Lett.",
  year: "2013"
},
subject: [...],
physical description: {...},
number of citations: 3,
other report number: {...},
title: {
  title: "Quintessential power-law cosmology: dark energy equation of state"
}

```

รูปที่ 3.57 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ Inspirehep

3.2.9 DBLP

DBLP เป็นฐานข้อมูลบรรณานุกรมทางด้านวิทยาการคอมพิวเตอร์ ซึ่งการดึงข้อมูลผ่าน API ของเว็บไซต์ DBLP นั้นจะต้องทำการแปลงชื่อนักวิจัยให้อยู่ในรูปแบบตามที่เว็บไซต์กำหนด ทำการดึงจะได้รายการหมายเลขผลงานของนักวิจัย จากนั้นนำหมายเลขผลงานของนักวิจัยมาทำการดึงข้อมูลต่อจึงจะได้ข้อมูลผลงานของนักวิจัย ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ DBLP มีดังนี้

- 1) ใส่ URL: <http://dblp.uni-trier.de/xml/docu/dblpxmlreq.pdf> ที่เว็บไซต์ เบราวเซอร์ ปรากฏหน้าวิธีการ และตัวอย่างการดึงข้อมูลผ่าน API ของเว็บไซต์ DBLP [19] ตามรูปที่ 3.58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

DBLP XML Requests

Appendix to the paper "DBLP — Some Lessons Learned" (June 17, 2009)

Michael Ley
Universität Trier, Informatik
D-54286 Trier
Germany
ley@uni-trier.de

ABSTRACT
If your software needs only in a few facts from DBLP, downloading the entire `dblp.xml` file may be a too costly burden. The web pages are intended for humans, wrappers are always exposed to the risk of formatting changes. In this appendix we describe a very basic API for DBLP. Our example application for the API is a simple crawler which finds the shortest path between two DBLP authors in the coauthor graph. In addition the appendix lists code to map person names to DBLP URLs.

DBLP Records
If you know the key of a DBLP record, you may retrieve the record from the URL

is used by the other services described in the sequel of this section.

Person Search
On <http://dblp.uni-trier.de/> DBLP provides a primitive form to search for persons inside the bibliography. If you type in `Schek`, you will get an answer on page <http://dblp.uni-trier.de/search/author?author=Schek>

This page is an HTML document. Modify the URL by inserting an `'x'` after the question mark:
.../search/author?xauthor=Schek

รูปที่ 3.58 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ DBLP

- 2) จากนั้นทำการแทน ชื่อ-นามสกุล ของนักวิจัยลงใน URL ที่ใช้แปลงชื่อผ่าน API ของเว็บไซต์ DBLP ตามรูปที่ 3.59 โดยใช้ชื่อ "Wasan Pattara-atikom" เป็นตัวอย่างในการค้นหาข้อมูล

URL = <http://dblp.uni-trier.de/search/author?xauthor=Wasan+Pattara-atikom>

รูปที่ 3.59 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ DBLP

- 3) จะได้ "urlpt" (ชื่อนักวิจัยตามรูปแบบที่เว็บไซต์กำหนด) ในรูปแบบของ XML ดังรูปที่ 3.60

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<authors>
  <author urlpt="p/Pattara-Atikom:Wasan" Wasan Pattara-Atikom</author>
</authors>
```

รูปที่ 3.60 ชื่อนักวิจัยตามรูปแบบที่เว็บไซต์กำหนดในรูปแบบ XML ของเว็บไซต์ DBL

- 4) นำค่าใน "urlpt" มาใส่ URL ที่ใช้ดึงรายการไอดีผลงานของนักวิจัยผ่าน API ของเว็บไซต์ DBLP ตามรูปที่ 3.61

URL = <http://dblp.uni-trier.de/rec/pers/p/Pattara=Atikom:Wasan/xk>

รูปที่ 3.61 ตัวอย่าง URL ที่ใช้ดึงรายการไอดีผลงานของนักวิจัยผ่าน API ของเว็บไซต์ DBLP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 5) ข้อมูลที่ดึงได้จะเป็นรายการหมายเลขผลงานของนักวิจัย (dblpkey) ในรูปแบบของ XML ตามรูปที่ 3.62

This XML file does not appear to have any style information associated with it. The

```

<dblpperson name="Wasan Pattara-Atikom">
  <dblpkey type="person record">homepages/29/1423</dblpkey>
  <dblpkey>conf/huc/WanichayapongSP15</dblpkey>
  <dblpkey>conf/jist/WanichayapongPP14</dblpkey>
  <dblpkey>conf/hicss/SaiprasertP13</dblpkey>
  <dblpkey>journals/comsur/PanichpapiboonP12</dblpkey>
  <dblpkey>conf/srii/SaikaewJJASPT12</dblpkey>
  <dblpkey>conf/srii/PrathombutrPC12</dblpkey>
  <dblpkey>journals/tvt/PanichpapiboonP11</dblpkey>
  <dblpkey>conf/medes/Pattara-AtikomA09</dblpkey>
  <dblpkey>journals/tvt/PanichpapiboonP08</dblpkey>
  <dblpkey>conf/vtc/PuntumaponP08</dblpkey>
  <dblpkey>conf/mswim/Pattara-AtikomBK04</dblpkey>
  <dblpkey>journals/sigmobile/Pattara-AtikomBK03</dblpkey>
  <dblpkey>journals/wc/Pattara-AtikomKB03</dblpkey>
  <dblpkey>conf/globecom/Pattara-AtikomKB03</dblpkey>
  <dblpkey>conf/icccn/Pattara-AtikomBK03</dblpkey>
</dblpperson>

```

รูปที่ 3.62 รายการไอดีผลงานของนักวิจัย (dblpkey) ในรูปแบบของ XML ของเว็บไซต์ DBLP

- 6) นำค่าใน “dblpkey” มาใส่ URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ DBLP ตามรูปที่ 3.63

URL = <http://dblp.uni-trier.de/rec/bibtex/journals/wc/Pattara-AtikomKB03.xml>

รูปที่ 3.63 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานของนักวิจัยผ่าน API ของเว็บไซต์ DBLP

- 7) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบ XML ตามรูปที่ 3.64 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ author, title, year, journal, url

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

<dblp>
  <article key="journals/wc/Pattara-AtikomKB03" mdate="2013-08-26">
    <author>Wasan Pattara-Atikom</author>
    <author>Prashant Krishnamurthy</author>
    <author>Sujata Banerjee</author>
    <title>
      Distributed mechanisms for quality of service in wireless LANs.
    </title>
    <pages>26-34</pages>
    <year>2003</year>
    <volume>10</volume>
    <journal>IEEE Wireless Commun.</journal>
    <number>3</number>
    <see>http://dx.doi.org/10.1109/MWC.2003.1209593</see>
    <url>db/journals/wc/wc10.html#Pattara-AtikomKB03</url>
  </article>

```

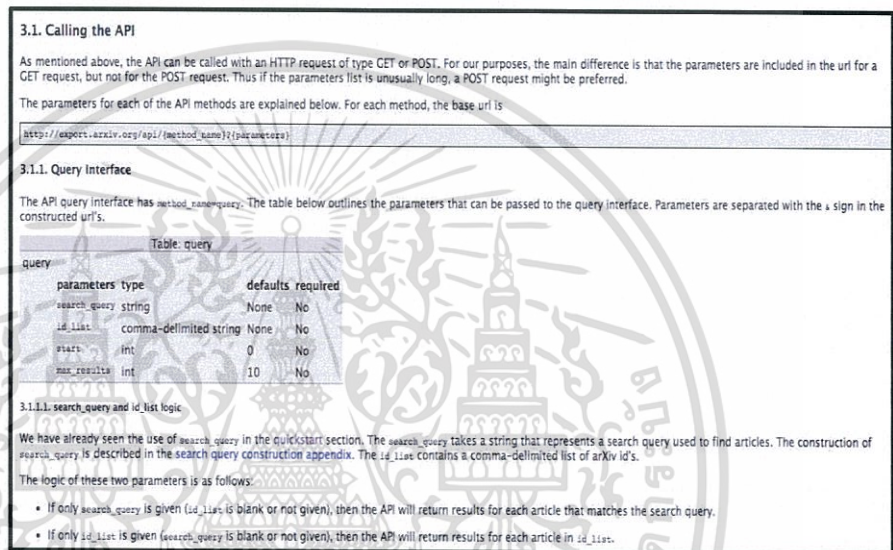
รูปที่ 3.64 ข้อมูลผลงานในรูปแบบ JSON ที่ดึงผ่าน API ของเว็บไซต์ DBLP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้拿去ใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.10 Arxiv

Arxiv เป็นเว็บไซต์สำหรับรวบรวมผลงานที่ตีพิมพ์และกำลังจะตีพิมพ์ ในสาขาฟิสิกส์ ดาราศาสตร์ คอมพิวเตอร์ ชีววิทยาเชิงปริมาณ และสถิติ ซึ่งขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ Arxiv มีดังนี้

- 1) ใส่ URL: <http://arxiv.org/help/api/user-manual> ที่เว็บเบราว์เซอร์ ปรากฏหน้าวิธีการ และ ตัวอย่างการดึงข้อมูลผ่าน API ของเว็บไซต์ Arxiv [20] ตามรูปที่ 3.65



3.1. Calling the API

As mentioned above, the API can be called with an HTTP request of type GET or POST. For our purposes, the main difference is that the parameters are included in the url for a GET request, but not for the POST request. Thus if the parameters list is unusually long, a POST request might be preferred.

The parameters for each of the API methods are explained below. For each method, the base url is

```
http://export.arxiv.org/api/method_name?(parameters)
```

3.1.1. Query Interface

The API query interface has method_name=query. The table below outlines the parameters that can be passed to the query interface. Parameters are separated with the & sign in the constructed url's.

parameters	type	defaults	required
search_query	string	None	No
id_list	comma-delimited string	None	No
start	int	0	No
max_results	int	10	No

3.1.1.1. search_query and id_list logic

We have already seen the use of search_query in the quickstart section. The search_query takes a string that represents a search query used to find articles. The construction of search_query is described in the search query construction appendix. The id_list contains a comma-delimited list of arxiv id's.

The logic of these two parameters is as follows:

- If only search_query is given (id_list is blank or not given), then the API will return results for each article that matches the search query.
- If only id_list is given (search_query is blank or not given), then the API will return results for each article in id_list.

รูปที่ 3.65 หน้าเว็บไซต์แสดงการวิธีการดึงข้อมูลผ่าน API ของเว็บไซต์ Arxiv

- 2) จากนั้นทำการแทน ชื่อ-นามสกุล ของนักวิจัยลงใน URL ที่ใช้ดึงข้อมูลผ่าน API ของเว็บไซต์ Inspirehep ตามรูปที่ 3.66 โดยใช้ชื่อ “Prasart Suebka” เป็นตัวอย่างในการค้นหาข้อมูล

URL = http://export.arxiv.org/api/query?search_query=author:Prasart+Suebka

รูปที่ 3.66 ตัวอย่าง URL ที่ใช้ดึงข้อมูลผลงานผ่าน API ของเว็บไซต์ Arxiv

- 3) ข้อมูลที่ดึงได้จะอยู่ในรูปแบบของ XML ตามรูปที่ 3.67 โดยผลลัพธ์ข้อมูลที่ทำการดึงได้แก่ title, author, published_date, link_abs, link_pdf

```

<entry>
  <id>http://arxiv.org/abs/hep-ph/0411238v1</id>
  <updated>2004-11-18T08:40:22</updated>
  <published>2004-11-18T08:40:22</published>
  <title>Photon number emission in synchrotron radiation: Systematics for
high-energy particles</title>
  <summary> A recent derivation of an explicit elementary expression for the mean number
&lt;N&gt; of photons emitted per revolution in synchrotron radiation allows a
systematic high-energy analysis leading to the result &lt;N&gt;=simeq
5\pi\alpha/\sqrt{3(1-\beta^2)}+a_0\alpha+\pi\alpha/\sqrt{1-\beta^2}/10\sqrt{3}
where $a_0$ is a constant, with relative errors of 2.2%, 0.64%, 0.017%, in
comparison to the well-known formula tabulated in the literature of 160%, 82%,
17% for $\beta=0.8, 0.9, 0.99$, respectively.
</summary>
  <author>
    <name>E. B. Manoukian</name>
  </author>
  <author>
    <name>N. Jearnkulprasert</name>
  </author>
  <author>
    <name>P. Suebka</name>
  </author>
  <arxiv:doi xmlns:arxiv="http://arxiv.org/schemas/atom">10.1393/ncb/i2002-10011-4</arxiv:doi>
  <link title="doi" href="http://dx.doi.org/10.1393/ncb/i2002-10011-4" rel="related"/>
  <arxiv:comment xmlns:arxiv="http://arxiv.org/schemas/atom">7 pages, no figures, LaTeX with cimento.cls</arxiv:comment>

```

รูปที่ 3.67 ข้อมูลผลงานในรูปแบบ XML ที่ดึงผ่าน API ของเว็บไซต์ Arxiv

3.3 แปลงข้อมูลจาก XML/JSON ให้อยู่ในรูปแบบที่ต้องการ

จากขั้นตอนการดึงข้อมูลผ่าน API ของเว็บไซต์ต่างๆ ข้อมูลที่ดึงมาได้จะอยู่ในรูปแบบ XML และ JSON ซึ่งต้องทำการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปใช้งานต่อได้ ซึ่งในแต่ละรูปแบบก็จะมีวิธีการแปลงข้อมูลแตกต่างกันออกไป

3.3.1 ขั้นตอนการแปลงข้อมูล XML

- 1) สมมติว่าเรามีไฟล์ XML ที่เรียกว่า "books.xml" ที่มีลักษณะตามรูปที่ 3.68

```

<?xml version="1.0" encoding="utf-8"?>
<bookstore>
  <book category="COOKING">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>
  <book category="CHILDREN">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="WEB">
    <title lang="en-us">XQuery Kick Start</title>
    <author>James McGovern</author>
    <year>2003</year>
    <price>49.99</price>
  </book>
  <book category="WEB">
    <title lang="en-us">Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>

```

รูปที่ 3.68 ตัวอย่างไฟล์ XML เรียกว่า "books.xml"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 2) แปลงข้อมูล XML โดยใช้คำสั่ง “simplexml_load_file("books.xml")” ตามรูปที่ 3.69

```
<?php
$xml=simplexml_load_file("books.xml") or die("Error: Cannot create object");
foreach($xml->children() as $books) {
    echo $books->title . ", ";
    echo $books->author . ", ";
    echo $books->year . ", ";
    echo $books->price . "<br>";
}
?>
```

รูปที่ 3.69 ตัวอย่างคำสั่งในการแปลงข้อมูล XML

- 3) จะได้ข้อมูลที่ได้อยู่ในรูปแบบทั่ว ตามรูปที่ 3.70

```
Everyday Italian, Giada De Laurentiis, 2005, 30.00
Harry Potter, J K. Rowling, 2005, 29.99
XQuery Kick Start, James McGovern, 2003, 49.99
Learning XML, Erik T. Ray, 2003, 39.95
```

รูปที่ 3.70 ตัวอย่างข้อมูลที่ได้จากการแปลงข้อมูล XML

3.3.2 ขั้นตอนการแปลงข้อมูล JSON

- 1) สมมติว่าเรามีไฟล์ JSON ที่เรียกว่า "employees.json" ที่มีลักษณะตามรูปที่ 3.71

```
{"employees": [
    {"firstName": "John", "lastName": "Doe"},
    {"firstName": "Anna", "lastName": "Smith"},
    {"firstName": "Peter", "lastName": "Jones"}
]}
```

รูปที่ 3.71 ตัวอย่างไฟล์ JSON ที่เรียกว่า "employees.json"

- 2) แปลงข้อมูล XML โดยใช้คำสั่ง “json_decode("employees.json")” ตามรูปที่ 3.72

```
<?php
    $json = json_decode( "employees.json");
    foreach($json->employees as $employee)
    {
        echo $employee->firstName . " ";
        echo $employee->lastName . "\n";
    }
?>
```

รูปที่ 3.72 ตัวอย่างคำสั่งในการแปลงข้อมูล JSON

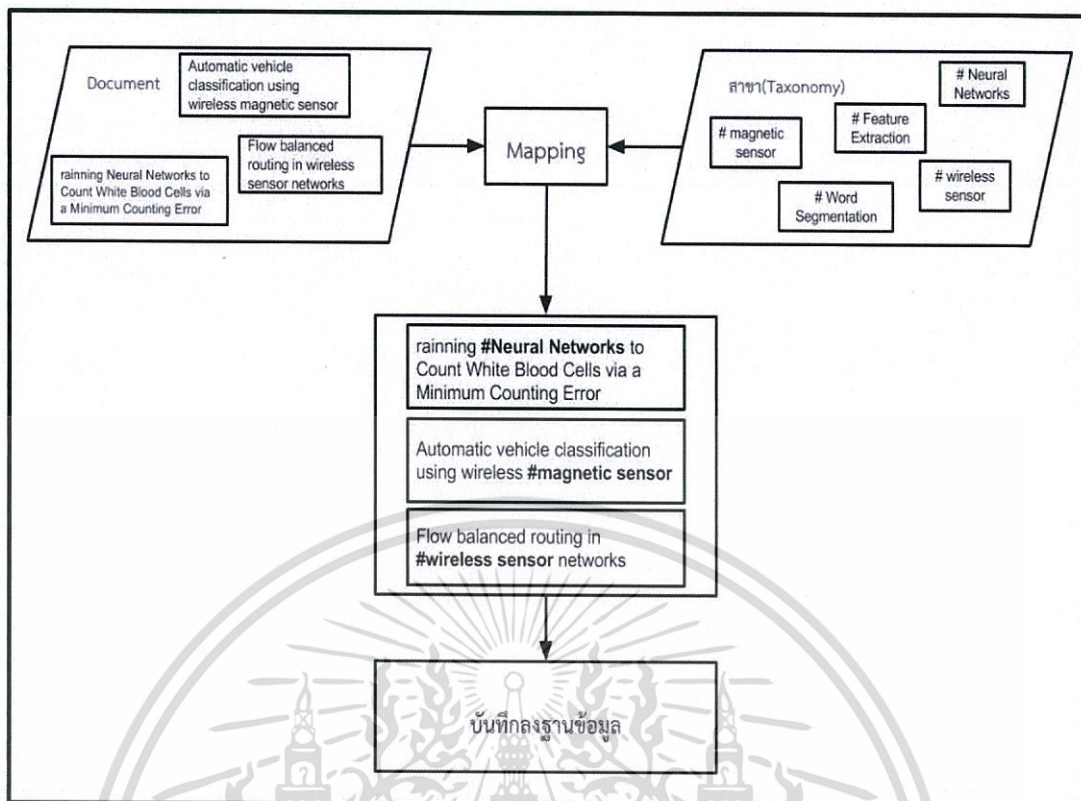
- 3) จะได้ข้อมูลที่ได้อยู่ในรูปแบบทั่ว ตามรูปที่ 3.73

```
John Doe
Anna Smith
Peter Jones
```

รูปที่ 3.73 ตัวอย่างข้อมูลที่ได้จากการแปลงข้อมูล JSON

3.4 ทำการ Mapping ข้อมูลผลงานกับสาขา

จากรูปที่ 3.74 เป็นขั้นตอนในการตัดแยกข้อมูลผลงานที่ได้จากการดึงข้อมูลผ่านผู้ให้บริการ API โดยการตัดแยกข้อมูลจะใช้วิธีการค้นหาสาขาต่างๆในฐานข้อมูลเปรียบเทียบกับรายละเอียดของผลงานวิจัย (Title or Content) กรณีที่ข้อมูลผลงานวิจัยตรงกับสาขาที่อยู่ในฐานข้อมูล ระบบจะทำการบันทึกผลการตัดแยกดังกล่าวลงในฐานข้อมูลเดิม



รูปที่ 3.74 ขั้นตอนการ Mapping ข้อมูลผลงานกับสาขา

3.5 เก็บผลงานข้อมูล

กระบวนการในการรวบรวมผลงานนักวิจัยจาก ฐานข้อมูลออนไลน์แบบอัตโนมัติ หลังจากที่ได้ข้อมูลต่างมาเรียบร้อยแล้ว ทำการบันทึกข้อมูลลงฐานข้อมูล โดยข้อมูลที่ดึงมาจากเว็บไซต์ต่างมีรายละเอียดตามรูปที่ 3.75

researcher_id	fname_th	lname_th	book_title	book_authors	book_publishedDate	keyword_en	keyword_th
4	ณัฐพงศ์	สว่างเมือง	A Location Fingerprint Framework Towards Efficient...	Nattapong Swangmuang	2011-09-04	EMS,Fingerprint,Indoor positioning,Location finger...	
6	สุทธิชัย	เปรมฤดี บริษัชาชาญ	Modeling of Photovoltaic Grid Connected Inverters ...	Alexis Polycarpou Belkacem Mahdad,C. M. Liaw,Dorin...	2011	Analysis,Grid connected,Identification,Inverters,M...	
8	มัลลิกา	บุญมี	Strategies for Lactate Removal for the Improved Pr...	Mallika Boonmee	2003	Lactococcus,Lactococcus lactis,Starter culture,US	Lactococcus,Lactococ lactis
12	อติศรา	เรืองแสง	การพัฒนากระบวนการผลิตผลงานชีวภาพจากถ้ำต้นปาดมอ...	สมพงษ์ โยทอง,อติศรา เรืองแสง	2010		ผลงานชีวภาพ,มีประสิทธิภาพ,ประสิทธิภาพผลิตผล...

รูปที่ 3.75 ตัวอย่างข้อมูลที่ทำการบันทึกผลงานข้อมูล

โดยจำนวนผลงานวิจัยที่ทำการดึงมาได้จากเว็บไซต์ทางวิชาการต่างๆ มีรายละเอียดจำนวนผลงาน ตามตารางที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 จำนวนผลงานที่ทำการรวบรวมข้อมูลจากเว็บไซต์ต่างๆ

รายชื่อเว็บไซต์	จำนวนผลงาน
PubMed	16,177
Slide Share	111
Google News Search	1,294
Google Books APIs	4,077
Google Video Search	1,885
Inspire-Hep	141
Arxiv	35
DBLP	2,715

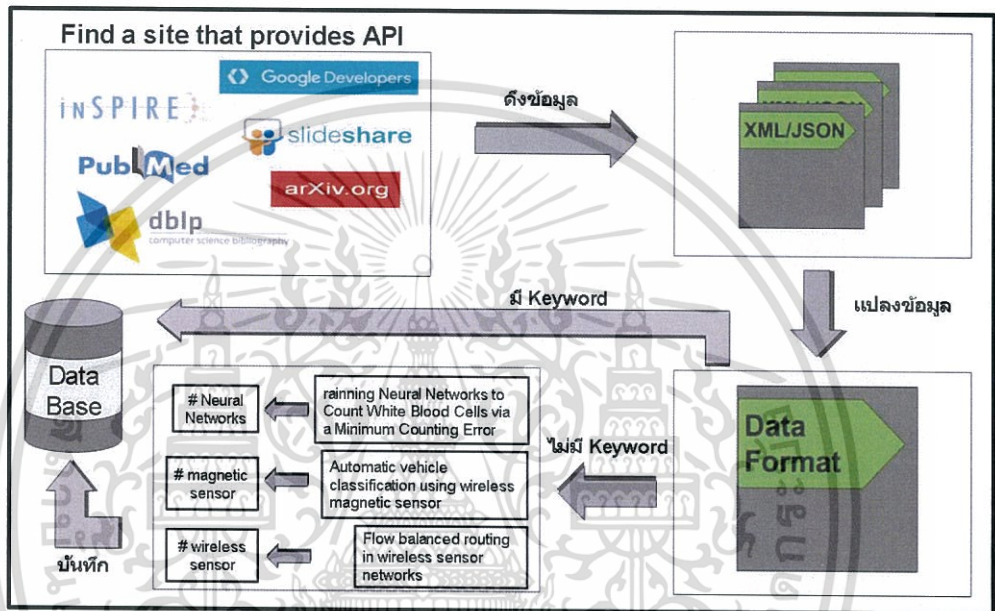


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและการอภิปราย

สำหรับกระบวนการรวบรวมผลงานนักวิจัยจากฐานข้อมูลออนไลน์แบบอัตโนมัติ ทำให้สามารถรวบรวมผลงานของนักวิจัยในแต่ละสาขาได้สะดวก และรวดเร็วมากยิ่งขึ้น โดยรูปที่ 4.1 แสดงถึงภาพรวมของขั้นตอนการดำเนินงาน ซึ่งประกอบด้วยขั้นตอน 5 ขั้นตอน ได้แก่



รูปที่ 4.1 ภาพรวมของขั้นตอนการดำเนินงาน

- 1) สืบค้นเว็บไซต์ที่ให้บริการ API
จากการค้นหารายชื่อผู้ให้บริการ API โดยทำการศึกษารายละเอียดของการให้บริการ รวมถึงข้อจำกัดต่างๆของการดึงข้อมูลผ่านผู้ให้บริการแต่ละราย
- 2) ดึงข้อมูลจากเว็บไซต์ผ่าน API
การดึงข้อมูลผ่านผู้ให้บริการ API จะมีวิธีการที่แตกต่างกันไปตามผู้ให้บริการแต่ละราย
- 3) แปลงข้อมูลจาก XML/JSON ให้อยู่ในรูปแบบที่ต้องการ
หลังจากทำการดึงข้อมูล ข้อมูลที่ได้จากผู้ให้บริการ API แต่ละรายจะมีการนำเข้าสู่กระบวนการแปลงรูปแบบ โดยทำการแปลงจากรูปแบบ XML/JSON เป็นรูปแบบที่สามารถนำไปใช้งานต่อได้
- 4) ทำการ Mapping ข้อมูลผลงานวิจัยกับสาขา
การ Mapping ข้อมูลระหว่างข้อมูลที่ได้จากการดึงผ่านผู้ให้บริการ API และทำการแปลงให้อยู่ในรูปแบบที่ต้องการ กับข้อมูลสาขาที่อยู่ภายในฐานข้อมูลจะใช้วิธีการค้นหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และเปรียบเทียบข้อมูลทั้งสองส่วน ในกรณีที่พบข้อมูลตรงกับสาขาที่อยู่ภายในฐานข้อมูล ระบบจะทำการบันทึกข้อมูลสาขาดังกล่าวกับข้อมูลผลงานวิจัยลงฐานข้อมูล

5) เก็บข้อมูลที่ได้ลงฐานข้อมูล (data dictionary)

จากกระบวนการดังกล่าวข้างต้น เมื่อได้ทำการ Mapping ข้อมูลระหว่างข้อมูล ผลงานวิจัย และสาขาที่อยู่ภายในฐานข้อมูล ระบบจะทำการเก็บผลการ Mapping ดังกล่าวลงในฐานข้อมูลเดิม โดยทำการตรวจสอบข้อมูลก่อนการบันทึกเพื่อลดความซ้ำซ้อน

ตารางที่ 4.1 จำนวนผลงาน และนักวิจัยที่มีผลงานจากเว็บไซต์ต่างๆ

	Slide Share	Google News	Google Video	Google books	Pubmed	Inspire	DBLP	Arxiv
จำนวนผลงาน	111	1,294	1,885	4,077	16,177	141	2,715	35
จำนวนนักวิจัย	65	712	888	2,509	3,103	29	699	13

จากตารางที่ 4.1 แสดงผลการดึงข้อมูลผลงานของนักวิจัย ซึ่งได้จากกระบวนการรวบรวมผลงานนักวิจัยจากแหล่งข้อมูลออนไลน์ผ่านเว็บไซต์ต่างๆ โดยพบข้อมูลผลงานจากเว็บไซต์ Pubmed มากที่สุด โดย Pubmed เป็นฐานข้อมูลที่ครอบคลุมวารสารสาขาชีววิทยาการแพทย์ซึ่งจากการทำการดึงข้อมูลพบผลงานของนักวิจัยจำนวน 16,177 ผลงานจากผลงานนักวิจัยทั้งหมด 26,435 ผลงาน และนักวิจัยที่มีผลงานบนเว็บไซต์ Pubmed มีจำนวนทั้งหมด 3,103 คน

ตารางที่ 4.2 จำนวนผลงาน และนักวิจัยที่มีผลงานตามประเภทผลงาน

	Presentation	News	Video	Book	Publication
จำนวนผลงาน	111	1,294	1,885	4,077	19,068
จำนวนนักวิจัย	65	712	888	2,509	3,844

จากตารางที่ 4.2 แสดงผลงานและนักวิจัยที่มีผลงานตามประเภทผลงาน ซึ่งได้จากกระบวนการรวบรวมผลงานนักวิจัยจากแหล่งข้อมูลออนไลน์ผ่านเว็บไซต์ต่างๆ โดยพบข้อมูลผลงานประเภท ผลงานตีพิมพ์ (Publication) มีจำนวนมากที่สุด โดยพบผลงานวิจัย 19,068 จากผลงานนักวิจัยทั้งหมด 26,435 ผลงาน และพบนักวิจัยที่มีผลงานตีพิมพ์ (Publication) ทั้งหมด 3,844 คน เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอขั้นตอนวิธีการรวบรวมข้อมูลผลงานวิจัยด้านวิทยาศาสตร์ จากแหล่งข้อมูลออนไลน์ โดยจะอธิบายวิธีการดึงข้อมูลผลงานของนักวิจัยผ่าน API จากเว็บไซต์ทางวิชาการต่างๆ ที่ให้บริการการดึงข้อมูลผ่าน API ซึ่งผลการดึงข้อมูลผลงานของนักวิจัยนั้น ผลงานที่ดึงได้ส่วนใหญ่มีข้อมูลที่ต้องการ ถึงแม้ว่าบางส่วนอาจจะไม่ค่อยตรงกับชื่อของนักวิจัยอยู่บ้างแต่ก็ถือว่าเป็นส่วนน้อย น่าจะช่วยเสริมให้ระบบการจัดลำดับความเชี่ยวชาญของนักวิจัยจากข้อมูลผลงาน สามารถประเมินความเชี่ยวชาญของนักวิจัยได้ดียิ่งขึ้น อีกทั้งผู้ที่สนใจยังสามารถใช้เป็นแนวทางในการศึกษา วิธีการดึงข้อมูลผ่าน API จากเว็บไซต์ต่างๆ ได้

โดยผลการวิจัย ซึ่งได้จากกระบวนการรวบรวมผลงานนักวิจัยจากแหล่งข้อมูลออนไลน์ผ่านเว็บไซต์ต่างๆ โดยพบข้อมูลผลงานจากเว็บไซต์ Pubmed มากที่สุด โดย Pubmed เป็นฐานข้อมูลที่ครอบคลุมวารสารสาขาชีววิทยาการแพทย์ซึ่งจากการทำการดึงข้อมูลพบผลงานของนักวิจัยจำนวน 16,177 ผลงาน จากผลงานนักวิจัยทั้งหมด 26,435 ผลงาน และนักวิจัยที่มีผลงานบนเว็บไซต์ Pubmed มีจำนวนทั้งหมด 3,103 คน นอกจากนี้จากกระบวนการรวบรวมผลงานนักวิจัยจากแหล่งข้อมูลออนไลน์ พบว่าผลงานวิจัยประเภทผลงานตีพิมพ์ (Publication) มีจำนวนมากที่สุด โดยพบผลงานวิจัย 19,068 จากผลงานนักวิจัยทั้งหมด 26,435 ผลงาน และพบนักวิจัยที่มีผลงานตีพิมพ์ (Publication) ทั้งหมด 3,844 คน

5.2 ข้อเสนอแนะ

- 1) ผู้พัฒนาจำเป็นต้องมีความอดทนในการดึงข้อมูล เนื่องจากการดึงข้อมูลไม่สามารถทำทั้งหมดในเวลาเดียวกันได้ ดังนั้นในการพัฒนาระบบจำเป็นต้องมีการหน่วงเวลาของโปรแกรมไว้ เพื่อป้องกันไม่ไห้ระบบของเว็บไซต์ที่ทำการดึงข้อมูล มองเป็นการโจมตีเซิร์ฟเวอร์ของทางเว็บไซต์ที่ได้ทำการดึงข้อมูลอยู่
- 2) ผู้ทำการพัฒนาระบบจำเป็นต้องศึกษาข้อกำหนดก่อนการดึงข้อมูลมาใช้งาน ซึ่งในบางเว็บไซต์อาจไม่ได้มีการระบุไว้
- 3) ผู้พัฒนาระบบควรทำการตรวจสอบความถูกต้องของฐานข้อมูลชื่อนักวิจัยที่จะใช้ในการดึงข้อมูลก่อนใช้งาน หากใช้ชื่อนักวิจัยไม่ถูกต้องจะทำให้ผลการค้นหาข้อมูลไม่ถูกต้องด้วยเช่นกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- [1] Mindphp. 2014. ประวัติความเป็นมาของภาษา PHP. [Online]. [Cited 2015 Nov 20].
Available: <http://www.mindphp.com/บทเรียนออนไลน์/บทเรียน-php/46-history-of-php.html>.
- [2] Mindphp. 2012. SQL คืออะไร. [Online]. [Cited 2015 Nov 20].
Available: <http://www.mindphp.com/คู่มือ/73-คืออะไร/2088-sql-คืออะไร.html>.
- [3] Wikipedia. 2015. XML. [Online]. [Cited 2015 Nov 24].
Available: <https://th.wikipedia.org/wiki/เอกซ์เอ็มแอล>.
- [4] Json. 2013. Introducing JSON. [Online]. [Cited 2015 Nov 24].
Available: <http://www.json.org>.
- [5] W3schools. 2015. JSON Tutorial. [Online]. [Cited 2015 Nov 24].
Available: <http://www.w3schools.com/json>.
- [6] ไกรศักดิ์ เกสร. 2011. “การค้นหาข้อมูลเชิงความหมาย.” วไลยอลงกรณ์ปริทัศน์. 1(2): 4
- [7] บุญเลิศ อรุณพิบูลย์. Text Mining. [Online]. [Cited 2015 Nov 25].
Available: <http://www.slideshare.net/boonlert/text-mining-data-mining>.
- [8] B. Rhode, T. Starner. 1996. “Remembrance Agent: A continuously running automated information retrieval system.” Technical Report, Media Lab, MIT
- [9] โชคชัย พุฒตาล. 2012. “การจัดลำดับความเชี่ยวชาญสำหรับสมาชิกกลุ่มเครือข่ายสังคมออนไลน์.”
วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ บัณฑิตวิทยาลัย.
จุฬาลงกรณ์มหาวิทยาลัย.
- [10] IEEE Xplore. 2010. Querying the IEEE Xplore Search Gateway. [Online]. [Cited 2015 Nov 3].
Available: <http://ieeexplore.ieee.org/gateway>.
- [11] ELSEVIER Developers. 2015. Elsevier Scopus APIs. [Online]. [Cited 2015 Nov 9].
Available: http://dev.elsevier.com/sc_apis.html.
- [12] Alexander Hadik. 2015. CREATE PUBMED CITATIONS AUTOMATICALLY USING PUBMED API. [Online]. [Cited 2015 SEP 10].
Available: <http://www.alexhadik.com/blog/2014/6/12/create-pubmed-citations-automatically-using-pubmed-api>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง(ต่อ)

- [13] NCBI. 2008. **Entrez Programming Utilities Help**. [Online]. [Cited 2015 SEP 17].
Available: <http://www.ncbi.nlm.nih.gov/books/NBK25500>.
- [14] SlideShare. 2015. **SlideShare API**. [Online]. [Cited 2015 May 19].
Available: <http://www.slideshare.net/developers/documentation>.
- [15] Google Developer. 2010. **News Search**. [Online]. [Cited 2015 JUN 8].
Available: <https://developers.google.com/news-search/v1/jsondevguide>.
- [16] Google Developer. 2010. **Google Video Search API**. [Online]. [Cited 2015 JUN 11].
Available: <https://developers.google.com/video-search/v1/jsondevguide>.
- [17] Google Developer. 2010. **Google Books APIs**. [Online]. [Cited 2015 JUN 11].
Available: https://developers.google.com/books/docs/v1/getting_started?hl=th.
- [18] Inspirehep. 2014. **Inspirehep API**. [Online]. [Cited 2015 Oct 16].
Available: <http://inspirehep.net/info/hep/api?ln=en>.
- [19] Michael Ley. 2009. **DBLP XML Requests**. [Online]. [Cited 2015 Oct 13].
Available: <http://dblp.uni-trier.de/xml/docu/dblpxmlreq.pdf>.
- [20] Cornell Univer Library. 2012. **Arxiv API User's Manual**. [Online]. [Cited 2015 Oct 21].
Available: <http://arxiv.org/help/api/user-ma>